

ORIGINAL ARTICLE

Open Access

CT and MRI radiomic features of lung cancer (NSCLC): comparison and software consistency



Chandra Bortolotto^{1,2}, Alessandra Pinto^{1*}, Francesca Brero^{3,4}, Gaia Messana¹, Raffaella Fiamma Cabini^{4,5*}, Ian Postuma⁴, Agnese Robustelli Test^{3,4*}, Giulia Maria Stella^{6,7}, Giulia Galli^{6,7}, Manuel Mariani³, Silvia Figini⁸, Alessandro Lascialfari^{3,4}, Andrea Riccardo Filippi^{2,9}, Olivia Maria Bottinelli² and Lorenzo Preda^{1,2}

Abstract

Background Radiomics is a quantitative approach that allows the extraction of mineable data from medical images. Despite the growing clinical interest, radiomics studies are affected by variability stemming from analysis choices. We aimed to investigate the agreement between two open-source radiomics software for both contrast-enhanced computed tomography (CT) and contrast-enhanced magnetic resonance imaging (MRI) of lung cancers and to preliminarily evaluate the existence of radiomic features stable for both techniques.

Methods Contrast-enhanced CT and MRI images of 35 patients affected with non-small cell lung cancer (NSCLC) were manually segmented and preprocessed using three different methods. Sixty-six Image Biomarker Standardisation Initiative-compliant features common to the considered platforms, PyRadiomics and LIFEx, were extracted. The correlation among features with the same mathematical definition was analyzed by comparing PyRadiomics and LIFEx (at fixed imaging technique), and MRI with CT results (for the same software).

Results When assessing the agreement between LIFEx and PyRadiomics across the considered resampling, the maximum statistically significant correlations were observed to be 94% for CT features and 95% for MRI ones. When examining the correlation between features extracted from contrast-enhanced CT and MRI using the same software, higher significant correspondences were identified in 11% of features for both software.

Conclusions Considering NSCLC, (i) for both imaging techniques, LIFEx and PyRadiomics agreed on average for 90% of features, with MRI being more affected by resampling and (ii) CT and MRI contained mostly non-redundant information, but there are shape features and, more importantly, texture features that can be singled out by both techniques.

Relevance statement Identifying and selecting features that are stable cross-modalities may be one of the strategies to pave the way for radiomics clinical translation.

Key points

- More than 90% of LIFEx and PyRadiomics features contain the same information.
- Ten percent of features (shape, texture) are stable among contrast-enhanced CT and MRI.

Chandra Bortolotto and Francesca Brero contributed equally to this work.

*Correspondence:

Alessandra Pinto

alessandra.pinto01@universitadipavia.it

Raffaella Fiamma Cabini

raffaellafiamm.cabini01@universitadipavia.it

Agnese Robustelli Test

agnese.robustellitest01@universitadipavia.it

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

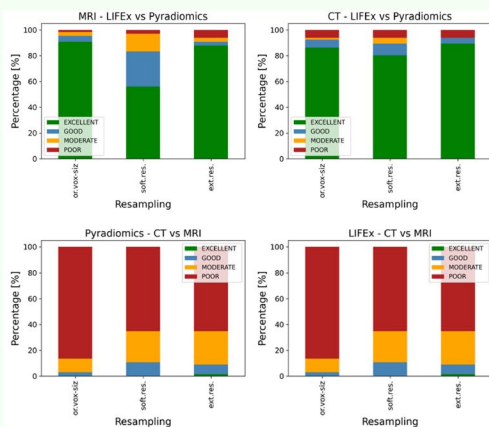
- Software compliance and cross-modalities stability features are impacted by the resampling method.

Keywords Biomarkers, Lung neoplasms, Magnetic resonance imaging, Radiomics, Tomography (x-ray computed)

Graphical Abstract

CT and MRI radiomic features of lung cancer: comparison and software consistency

- For both CT and MRI, LIFEx and Pyradiomics agree on average for 90% of features, with MRI more affected by resampling.
- CT and MRI contain mostly non-redundant information, but there are shape features and, more importantly, texture features that can be singled out by both techniques.



Agreement between LIFEx and PyRadiomics evaluated through intraclass correlation coefficient for MRI (left) and CT (right).

Comparison between CT and MRI features evaluated through intraclass correlation coefficient for Pyradiomics (left) and LIFEx (right).

Identifying and selecting features that are stable cross-modalities may be one of the strategies to pave the way for the radiomics clinical translation.



**Eur Radiol Exp (2024) Bartolotto C, Pinto A, Brero F, et al.
DOI: 10.1186/s41747-024-00468-8**

Background

Radiomics is a quantitative method that allows extracting mineable high-dimensional data, named radiomic features, from digital medical images [1–3]. The core hypothesis is that radiomic features might integrate visual analysis by unveiling tissue details and heterogeneity linked to image intensity distribution at an almost microscopic scale [4–6], thus becoming novel biomarkers. In oncology, radiomic features could, in principle, describe both the microenvironment [7] and the histotype and genotype of the tumor mass, supporting diagnosis determination, defining the prognosis, and predicting the therapeutic response [8]. The widespread interest in this method arises from several aspects. Radiomics, unlike biopsy, non-invasively captures information about the entire tumor [9]. Moreover, radiomics allows several evaluations of diagnostic images that are routinely acquired and permits to conduct analyses at different time points [10]. Radiomic features can be divided into four main categories: shape, histogram-based, texture, and wavelet [11]. In

general, shape ones refer to the geometric properties of a region of interest; first-order features are related to the image intensity histograms; texture features define the mathematical relationship of a single voxel with one or more neighboring voxels reflecting, *e.g.*, the intratumoral heterogeneity; and wavelet features are filter-based features able to enhance some characteristics of the image, analyzing its frequency domain information [12, 13].

Over the years, lung cancer has been the subject of several radiomic studies, being the second most frequent cancer and the leading cause of cancer-related death worldwide [14]. Many studies have been conducted by using computed tomography (CT) and positron emission tomography (PET) acquisitions, which are already widely used in the daily management of lung cancer [15–19]. Despite the great interest in integrating lung magnetic resonance imaging (MRI) into clinical practice, the lung remains one of the few anatomical sites in which MRI has not yet reached CT performances [20, 21]. Even though MRI does not expose

patients to ionizing radiation and provides optimal soft tissue contrast and exclusive morpho-functional information, it is still underused due to unfavorable occurrences [22].

The main obstacles to obtain good lung MRI images are related to the low signal-to-noise ratio (SNR) caused by lung parenchyma's poor proton density, frequent tissue-air interfaces, movement artifacts, and the lack of standard protocols [23]. In the literature, there are few studies concerning the extraction of radiomic features from MRI of lung tumors. For example, one study determined the optimal timing post-contrast injection to extract radiomic features on T1-weighted images for predicting 2 years of progression-free survival [24]. Another preliminary study suggested that MRI-derived radiomic features may improve the accuracy of models that predict therapy response and survival at different time points, compared to that of models based on CT features only [10]. Similarly, few studies have investigated the correlation between radiomic features extracted from different imaging techniques. Mahon et al. [10] have investigated the repeatability of texture features derived from CT and MRI of lung cancer. On the other hand, Vuong et al. [13] have analyzed the correlation between features extracted from PET/CT and PET/MRI images, finding a close correlation between them.

In this scenario, we aimed to conduct a preliminary methodological investigation, exploring the correlation between CT and MRI lung cancer radiomic features, to assess whether specific lung cancer intrinsic aspects can be depicted by both imaging modalities. Since CT and MRI are based on different physical principles, radiomic features extracted with these techniques are generally not directly comparable, even considering the lower-level ones. CT scans employ x-rays to produce detailed images of the body, which describe tissue's electronic density, while MRI uses magnetic fields and radiofrequency pulses to generate images that reflect tissue's complex properties such as proton density, nuclear relaxation times, and many other, including parameters related to functional behavior. Consequently, we hypothesize that if some features are directly comparable between CT and MRI, we could assume that they are strongly related to tissue/organ biology and physiology.

To evaluate the correlation between CT and MRI features, we employed two open-access software, LIFEx [25] and PyRadiomics [26], with a double purpose. The first objective was to evaluate the correlation between LIFEx and PyRadiomics features, chosen on a broad range, as reported also in literature [27, 28], for two different imaging modalities evaluated separately on the same patient cohort. The second objective was to establish possible correlations among CT and MRI radiomic features,

determining at the same time which software enables the extraction of such features. Lastly, we evaluated the impact of the voxel resampling algorithm on the two previous goals, as is known from the literature that the choice made at this stage can impact the analysis [29, 30].

Methods

Patients

This study was approved by the local Medical Research Ethics Committee of Fondazione IRCSS Policlinico San Matteo (Protocol code P_20130113422), and informed consent was obtained from all participants.

Thirty-five patients with non-small cell lung cancer (NSCLC), histologically confirmed from April 2021 to June 2023, were prospectively included as the study participants. The cohort consists of 26 males (74%) and 9 females (26%), with ages ranging from 49 to 84 years (median age 68 years). Concerning the NSCLC histotype distribution, 13 patients (37%) had an adenocarcinoma, 12 patients (34%) a squamocellular carcinoma, and the remaining 10 (29%) a poorly differentiated NSCLC. Tumor size was between 2 cm and a maximum of 15 cm with a corresponding stage between the II and the IV stage (in particular, 4 of II, 20 of III, and 11 of IV).

The following patients were excluded: (i) patients who had no adequate compliance capabilities and/or characteristics for undergoing MRI (e.g., claustrophobia, contraindications to MRI such as pacemakers, contraindications to Gd-based contrast agents); (ii) those who had received treatment before imaging; (iii) patients with lung tumors not classified as NSCLC.

Image acquisition

CT protocol

All patients underwent a thoracic CT examination, in a supine position from the apex to the base of the lung. Conventional CT was performed with a 64-slice scanner (SOMATOM Flash; Siemens Healthineers, Erlangen, Germany) for 16 patients, with a 16-slice scanner (SOMATOM Sensation; Siemens Healthineers) for 8 patients, with a 64-slice scanner (SOMATOM Sensation; Siemens Healthineers) for 6 patients, with a 160-slice scanner (Aquilion PRIME; Canon Medical Systems, Otawara, Japan) for 2 patients, and with a 320-slice CT (Aquilion ONE; Canon Medical Systems) for 3 patients. The scanning parameters were tube voltage 120 kV, tube current automatically modulated, slice thickness 2 mm, slice spacing 1 mm, pitch 1, rotation time 0.5 s, matrix 384×384, field of view set to 300 mm, and then adapted to patients. The scanning was completed under breath-hold condition and directly after the intravenous iodinated contrast medium injection (iomeprol 350 mgI/mL, 2 mL/s, 120 mL, 40-mL saline flush), in the venous phase

(i.e., 60–90 s after injection). After scanning, the original images were set to be the mediastinal window (smooth-medium kernel), automatically reconstructed.

MRI protocol

All patients underwent thoracic MRI in a supine position from the apex to the base of the lung. Conventional MRI was performed with a 1.5-T system (MAGNETOM Aera; Siemens Healthineers) using a 32-channel surface coil. During the examination, both free-breathing and breath-hold sequences were used. Scan sequences included in the study were axial and coronal volumetric interpolated breath-hold examination – VIBE T1-weighted sequences after the intravenous injection of paramagnetic contrast medium (gadoterate meglumine, 0.2 mL/kg (0.1 mmol/kg), 2 mL/s, 40-mL saline flush). For axial scanning, the matrix was 320×320, repetition time 2.1 ms, echo time 0.72 ms, field of view 450×350 mm, slice thickness 2.5 mm, layer spacing 0 mm, and number of layers 96.

Image selection

Chest axial CT images and axial MRI T1-weighted images after contrast medium injection were selected for feature extraction in this study. We opted to include only contrast-enhanced T1-weighted MRI images to facilitate a direct comparison with the portal-venous phase CT scans. This choice was driven by the need to select the MRI sequence that best resembles the CT phase from a visual radiological and pharmacokinetic point of view.

Tumor segmentation

Contrast-enhanced axial chest CT axial T1-weighted MRI in the Digital Imaging and Communications in Medicine – DICOM format were imported into the ITK-SNAP software (<http://www.itksnap.org>) and manually segmented.

The segmentation of CT images was made semiautomatically, using a Hounsfield unit seed-based method, while the segmentation of MRI was performed completely manually considering the lack of automatic or semiautomatic options. The segmentations were performed by three different radiologists (A.P., G.M., and C.B. with 2, 4, and 7 years of experience in thoracic imaging); complex cases were reviewed collegially with the aid of a fourth expert thoracic radiologist (L.P.). Cases were randomly assigned to each operator, and CT and MRI were not presented simultaneously; the minimal interval period between CT and MRI segmentation was 21 days. The region of interest of the tumor was segmented slice by slice to obtain the whole volume of interest by summing the segmented areas on each slice; the original images and the corresponding volume of interest image were saved using the

Neuroimaging Informatics Technology Initiative – NIfTI-1 format (<https://nifti.nimh.nih.gov/nifti-1>).

Image preprocessing

As suggested by previous literature [31–35] to reduce variability between images, it is necessary to preprocess the images. Regarding radiomic features, some image characteristics are more influential than others. In particular, the gray-level distribution and voxel size are highly relevant. Regarding the voxel resampling, three different strategies have been considered:

1. Features have been computed without performing resampling (original voxel size).
2. Features have been computed after setting resampling voxel dimensions directly on the software considered (software resampling).
3. All images have been resampled into the same voxel space using the Python package Nibabel (<https://github.com/nipy/nibabel>) before performing the radiomic features extraction (external resampling).

For the second and third resampling modalities, the voxel size has been set to $1 \times 1 \times 1 \text{ mm}^3$ for CT and $1.4 \times 1.4 \times 1.4 \text{ mm}^3$ for MRI.

We have also normalized the signal intensity distribution of MRI images through the histogram matching technique, as proposed in previous works [36]. Specifically, by using the SimpleITK Python library [37], we transformed the intensity histogram of the images to align with the histogram of a reference image from a healthy subject. As the final preprocess step, we have discretized the image gray level distribution in 64 bins.

Feature selection and extraction

We used the freeware LIFEx [25] version 7.3.0 and the open-source Python package PyRadiomics [26] version 3.0.1, Image Biomarker Standardisation Initiative (IBSI) [31] compliant. Both platforms allow to customize several parameters (spatial resampling, rescaling, and so on). The spatial resampling customization was performed just for the internal method resampling. The interpolator chosen is the sitkBSpline [38] and the intensity range was discretized in 64 bins, extracting features fixing the bin number, as advised when intensity units are arbitrary, as is for MRI [31]. We selected 66 IBSI-compliant features common to both software reported in Table 1. The extracted features were from six different feature categories: first-order features, shape features, and features from four different textures subdomains: gray-level co-occurrence matrix (GLCM), gray-level run length matrix

Table 1 List of all 66 IBSI-compliant features common to both software (LIFEx and PyRadiomics), divided into different classes

Shape	Histogram		GLCM	GLRLM	GLZLM	NGTDM	
Voxel Volume	Skewness	10th Percentile	Contrast	ClusterShade	Short Run	Small Zone	Coarseness
Surface Area	Kurtosis	90th Percentile	Correlation	DifferenceVariance	Emphasis (SRE)	Emphasis (SZE)	Complexity
Sphericity	Entropy	Maximum	Dissimilarity	Long Run	Long Run	Large Zone	Busyness
Maximum3DDiameter	Energy	InterquartileRange	Energy	DifferenceEntropy	Emphasis (LRE)	Emphasis (LZE)	Strength
	Uniformity	Range	Entropy	InverseVariance	Gray Level	Gray Level Non	Contrast
	Mean	MeanAbsolute Deviation	InverseDifference	SumEntropy	Non-Uniformity (GLNU)	Uniformity (GLNU)	
	Median	(MAD)	Autocorrelation	JointVariance	Run Length	Zone Size Non	
	Minimum	RobustMean	JointAverage	JointMaximum	Non-Uniformity (RLNU)	Uniformity (ZLNU)	
		AbsoluteDeviation(rMAD)	ClusterProminance	NormalizedInverse Difference (NID)	Run Percentage (RP)	Zone Percentage (ZP)	
		Variance	ClusterTendency		Low Gray Level Run Emphasis (LGRE)	Gray Level Variance (GLV)	
					High Gray Level Run Emphasis (HGRE)	Low Gray Level Zone Emphasis (LGZE)	
					Short Run Low Gray Level Emphasis (SRLGE)	High Gray Level Zone Emphasis (HGZE)	
					Short Run High Gray Level Emphasis (SRHGE)	Small Zone Low Gray Level Emphasis (SZLGE)	
					Long Run Low Gray Level Emphasis (LRLGE)	Small Zone High Gray Level Emphasis (SZHGE)	
					Long Run High Gray Level Emphasis (LRHGE)	Large Zone Low Gray Level Emphasis (LZLGE)	
						Large Zone High Gray Level Emphasis (LZHGE)	

GLCM Gray-level co-occurrence matrix, GLRLM Gray-level run length matrix, GLZLM Gray-level size zone matrix, NGTDM Neighboring gray-tone difference matrix

(GLRLM); gray-level size zone matrix (GLZLM); and neighboring gray-tone difference matrix (NGTDM) [8].

moderate ($0.5 \leq ICC < 0.75$), good ($0.75 \leq ICC < 0.9$), and excellent ($ICC \geq 0.9$) reliability.

Statistical analysis

All the statistical analysis has been conducted using Python (version 3.8.10, <http://www.python.org>). To assess the features’ agreement between the two radiomic software (for the same imaging modality) and between the two different imaging modalities (for the same software), the intraclass correlation coefficient (ICC) has been calculated (two-way mixed effects, absolute agreement single measurement configuration [39]). The ICC value was calculated as follows:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k - 1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where MS_R =mean square for rows; MS_E =mean square for error; MS_C =mean square for columns; n =number of subjects; and k =number of raters/measurements. We divided the ICC values into four ranges: poor ($ICC < 0.5$),

Results

Agreement between LIFEx and PyRadiomics software

The agreement between features computed from LIFEx and PyRadiomics was assessed for both CT and MRI and the three different voxel resampling strategies. The ICC values divided into four confidence levels for MRI (left) and CT (right) are shown in Fig. 1.

As regards MRI, excellent or good reliability was achieved by 95% of features without voxel resampling, 83% of features with the internal resampling software, and 91% of features with the external resampling software, as summarized in Table 2. Poor or moderate reliability was observed for 5% of features without any voxel resampling, 17% of features with the internal resampling software, and 9% of features with the external resampling software. For each of the resampling methods, moderate reliability was associated with MAX3DDiameter (SHAPE) and Sum Entropy (GLCM). Moreover, the

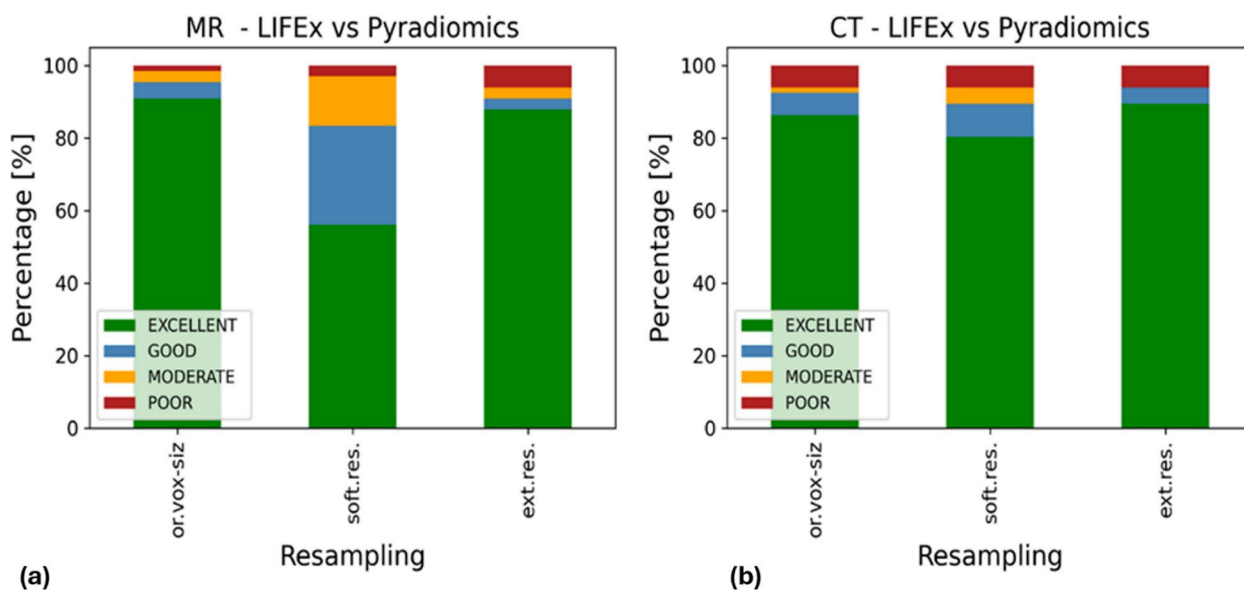


Fig. 1 Consistency between radiomic features extracted with LIFEx and PyRadiomics, expressed through ICC, from MRI (a) and CT (b) images. *Ext.res.* External resampling, *Or.vox-siz.* Original voxel size, *Soft.res.* Software resampling

Table 2 LIFEx versus PyRadiomics. Agreement between LIFEx and PyRadiomics for both MRI and CT, divided by ICC ranges. The results are presented in terms of radiomic features percentage exhibiting specific ICC values across the considered resampling methods

Intraclass correlation coefficient					
Imaging technique	Resampling	Excellent [%]	Good [%]	Moderate [%]	Poor [%]
MRI	Or.vox-siz	91.0	4.5	3.0	1.5
	Soft.res	56.1	27.3	13.6	3.0
	Ext.res	87.9	3.0	3.0	6.1
CT	Or.vox-siz	86.0	6.1	1.5	6.1
	Soft.res	80.3	9.1	4.5	6.1
	Ext.res	89.4	4.5	0	6.1

Ext.res External resampling, *Or.vox-siz.* Original voxel size, *Soft.res.* Software resampling

feature Inverse Variance (GLCM) exhibited poor reliability across all three resampling strategies.

Considering CT images, 92% of features demonstrated excellent or good reliability without voxel resampling, 89% of features had excellent or good reliability with internal resampling, and 94% of features exhibited excellent or good reliability with external resampling, as summarized in Table 2. Poor or moderate reliability was observed for 8% of features without voxel resampling, 11% of features with internal resampling, and 6% of features with external resampling. The three resampling strategies exhibited poor reliability for GLCM Inverse Variance, GLZLM LZE, GLZLM LZHGE, and GLZLM LZLGE. Figures 2 and 3 illustrate the variability of LIFEx-PyRadiomics features values for MRI and CT, respectively. These figures represent the distributions of a feature with excellent ICC and one with poor ICC.

Correlation between CT and MRI features

The comparison between features computed from CT and MRI was performed for the three voxel resampling strategies and the two radiomic software separately. Figure 4 summarizes the distribution of the ICC values in the four ranges of agreement. As regards PyRadiomics, only 3% of features had excellent or good reliability without voxel resampling, 11% of features showed excellent or good reliability with internal resampling, and 9% of features possessed excellent or good reliability with external resampling (Table 3). A poor or moderate agreement was obtained for 97% of features without voxel resampling, 89% of features with internal resampling, and 91% of features with external resampling. Table 4 provides a summary of the features with excellent and good reliability. Excellent or good agreement was observed between features extracted from different imaging modalities for

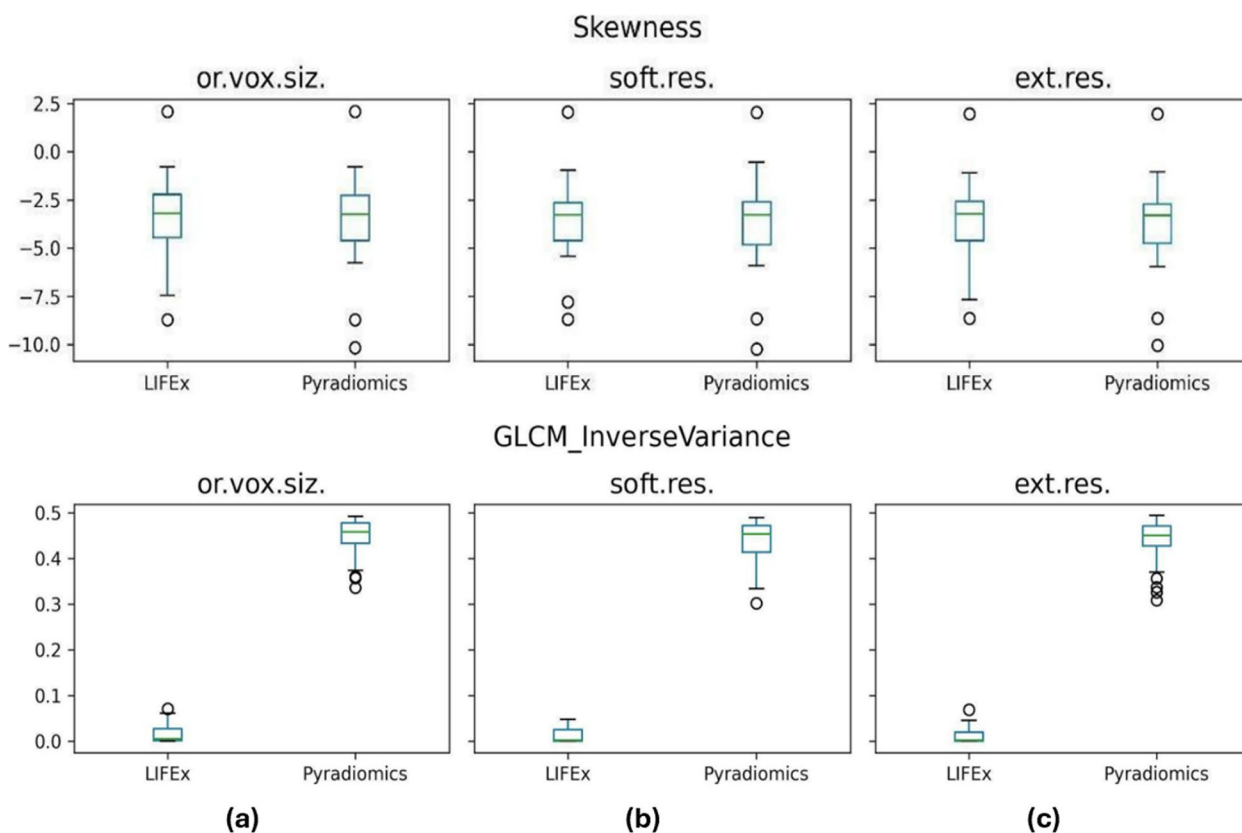


Fig. 2 Comparison between features distributions computed by LIFEx and PyRadiomics platforms from MRI acquisitions. The top panel line represents a feature with excellent reliability (Skewness), while the bottom panel line a feature with poor reliability (GLCM Inverse Variance) for the three options considered: original voxel size (a); software (internal) resampling (b); external resampling (c). *Ext.res.* External resampling, *Or.vox-siz.* Original voxel size, *Soft.res.* Software resampling

SHAPE features Volume and Surface Area, as well as for a few texture-based features. When considering all the analyzed resampling methods, the features demonstrating good/excellent reliability across all methods are the SHAPE ones. However, focusing on both internal and external resampling reveals additional common features, NGTDM Busyness, NGTDM Strength, and GLZLM ZP.

Considering LIFEx, reliability between imaging modalities was excellent or good for 5% of features without voxel resampling, 9% of features with internal resampling, and 11% of features with external resampling (Table 3). Reliability was poor and moderate for 95% of features without voxel resampling, 91% of features with internal resampling, and 89% of features with external resampling. Features with good agreement between imaging modalities belong to SHAPE features (Volume, Surface Area, and MAX3D Diameter) and texture features, as detailed in Table 4. Only the SHAPE feature Volume resulted in an excellent ICC for the external resampling. As found for PyRadiomics, texture features presented good ICC just for internal and external resampling. In this case, features

common to both resampling methods, with good ICC, are NGTDM Busyness and Strength, and GLRLM RLNU.

Discussion

The first part of our study evaluated the correlation between features extracted by LIFEx and PyRadiomics from contrast-enhanced CT and MRI of NSCLCs. Our purpose is to verify whether there is an agreement between the two radiomic software, considering two different imaging techniques.

For what concerns MRI, at least 83% of the features showed a good or excellent ICC for each of the considered resampling methods. The maximum agreement between LIFEx and PyRadiomics was obtained for images with original voxel dimensions (94.5% of the features), while the minimum agreement was gained by the internal resampling method (83% of the features). Possibly, the result found in this last case arises from the distinct resampling algorithms implemented in LIFEx and PyRadiomics. This effect could have been emphasized also by the upsampling operation made on the original images. In

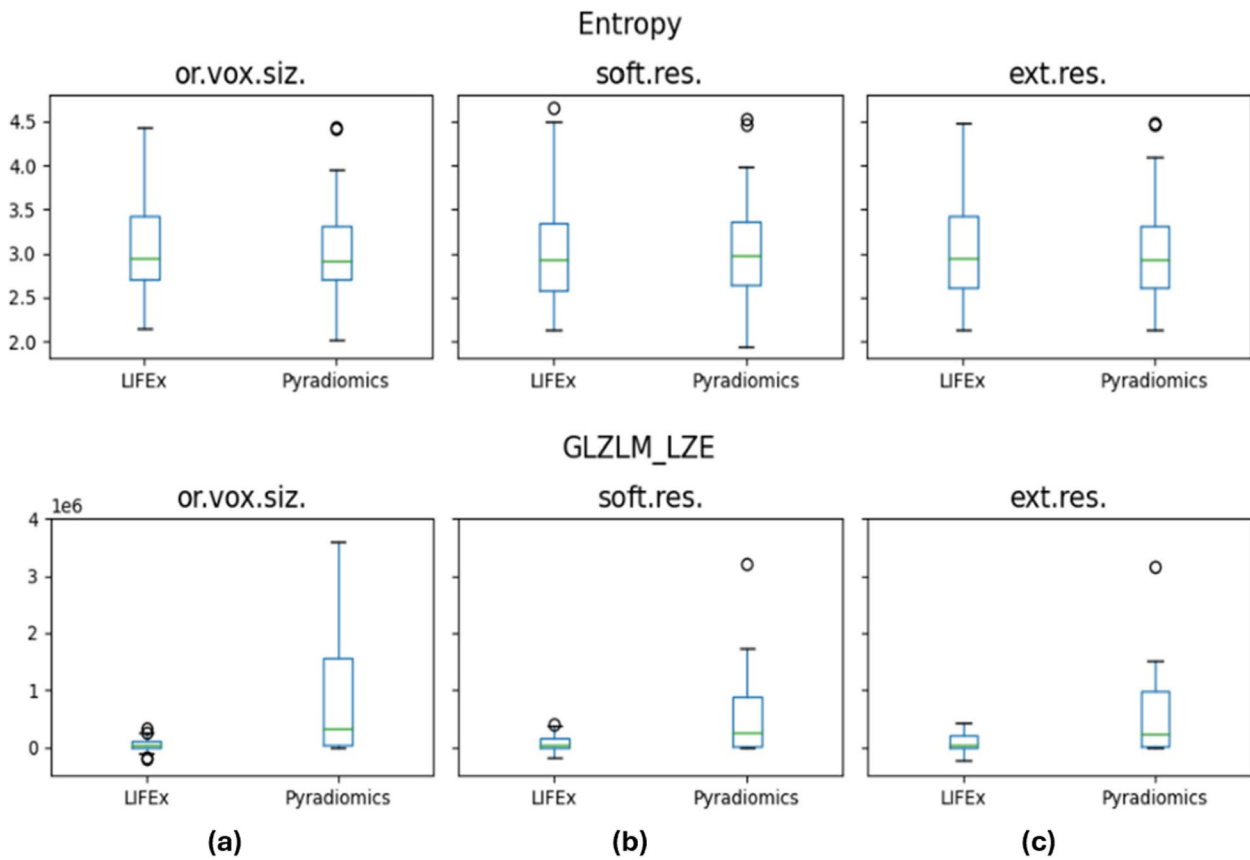


Fig. 3 Comparison between feature distributions computed by LIFEx and PyRadiomics platforms from CT images. The top panel line shows a feature with excellent reliability (Entropy), while the bottom panel line displays a feature with poor reliability (GLZLM Large Zone Emphasis), for the three options considered: original voxel size (a); software (internal) resampling (b); external resampling (c). *Ext.res.* External resampling, *Or.vox-siz.* Original voxel size, *Soft.res.* Software resampling

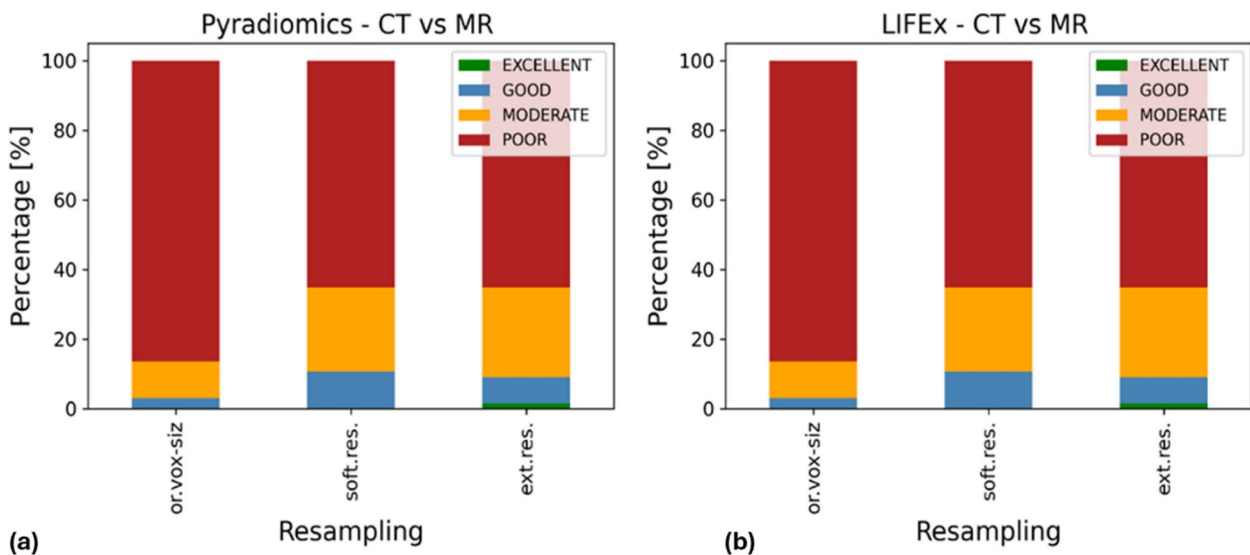


Fig. 4 Agreement between radiomic features extracted from CT and MRI images, expressed through ICC, for PyRadiomics (a) and LIFEx (b). *Ext.res.* External resampling, *Or.vox-siz.* Original voxel size, *Soft.res.* Software resampling

Table 3 CT versus MRI. CT-MRI concordance expressed through the percentage of radiomic features exhibiting a specific ICC, for both LIFEx and PyRadiomics. The results are reported for each of the considered resampling methods

Intraclass correlation coefficient					
Radiomic software	Resampling	Excellent [%]	Good [%]	Moderate [%]	Poor [%]
LIFEx	Or.vox-siz	0	4.5	7.5	88
	Soft.res	0	9.1	27.3	63.6
	Ext.res	1.5	9.1	22.7	66.7
PyRadiomics	Or.vox-siz	0	3.0	10.6	86.4
	Soft.res	0	10.6	24.2	65.2
	Ext.res	1.5	7.6	25.8	65.1

Ext.res. External resampling, Or.vox-siz. Original voxel size, Soft.res. Software resampling

Table 4 Radiomic features showing excellent and good ICC coefficients in CT-MRI comparison, considering PyRadiomics and LIFEx. For both software, the sole feature with excellent ICC is Voxel Volume (SHAPE), only employing external resampling

PyRadiomics software			LIFEx software		
Original voxel size	Internal resampling	External resampling	Original voxel size	Internal resampling	External resampling
SHAPE_Voxel Volume	SHAPE_Voxel Volume	SHAPE_Voxel Volume	SHAPE_Voxel Volume	SHAPE_Voxel Volume	SHAPE_Voxel Volume
SHAPE_Surface Area	SHAPE_Surface Area	SHAPE_Surface Area	SHAPE_Surface Area	SHAPE_Surface Area	SHAPE_Surface Area
	NGTDM_Busyness	NGTDM_Busyness	SHAPE_Maximum3D-Diameter	SHAPE_Maximum3D-Diameter	SHAPE_Maximum3D-Diameter
	NGTDM_Strength	NGTDM_Strength		NGTDM_Strength	NGTDM_Strength
	GLZLM_Zone_Percentage(ZP)	GLZLM_Zone_Percentage(ZP)		GLRLM_RLNU	GLRLM_RLNU
	GLCM_DifferenceEntropy	GLRLM_RLNU			GLZLM_Zone_Percentage(ZP)
	GLCM_NID				

GLCM Gray-level co-occurrence matrix, GLRLM Gray-level run length matrix, GLZLM Gray-level size zone matrix, NGTDM Neighboring gray-tone difference matrix, NID Normalized inverse difference, RLNU Run length non-uniformity

particular, the in-plane dimension was conserved, while the z-axis has been modified to obtain an isotropic voxel.

Considering CT images, we have found that at least 89% of features showed a good/excellent ICC for each of the considered resampling methods. The highest agreement between LIFEx and PyRadiomics was obtained for external resampling. The percentage of CT features exhibiting good/excellent reliability showed less variation among the resampling methods, compared to MRI. This could be addressed by the fact that CT images exhibit lower noise levels when compared to MRI ones. Moreover, the original voxel dimensions of CT images were closer to the isotropic voxel size than those of MRI, potentially reducing the impact of resampling on the distribution of extracted features. CT features that revealed moderate/poor reliability are the same for each of the resampling methods: GLCM Inverse Variance, GLZLM LZE, GLZLM LZHG, and GLZLM LZLGE.

Some considerations can be made for both CT and MRI. Firstly, the software agreement was higher when considering external resampling compared to internal resampling. This once again underlines the feature's dependency on the choices made in each step of the

radiomic workflow. Secondly, considering the high percentage of concordant features between LIFEx and PyRadiomics for both CT and MRI, it is possible to conclude that the two software agree with each other, regardless of the imaging technique used. These results were achieved following the IBSI guidelines and the conclusions obtained from previous works [27, 31]. It is noteworthy that LIFEx-PyRadiomics agreement was achieved considering two different imaging techniques. This is an important result, especially for lung MRI, given the challenges posed by artifacts, low signal, and other well-known limits.

The second part of the study investigated which of the two software packages could extract the highest number of correlated features among CT and MRI. Our purpose was to verify whether there is some lung cancer intrinsic information that can be depicted by both CT and MRI. We were not expecting a high number of features to be correlated between the two imaging modalities, as they rely on different physical principles, even though the local effect of iodine and Gd-based agents is an increase in x-ray attenuation (CT scans) or signal (T1-weighted MRI), which always translates in an increase of the values

of the gray scale in the images. We aimed to address the question regarding the possible existence of information strictly linked to NSCLC biology, beyond the considered imaging technique.

From the medical point of view, it should be noticed that the low percentage (10%) of correlated radiomic features between MRI and CT can be considered informative about lung cancer. Features stable across modalities may carry relevant biological information, showing the ability to reflect histopathological phenomena, such as inflammation or vascularization, related to lung cancer's characterization and options for treatment. Thus, with higher statistics of patients, selecting features that are cross-modality and stable (even few) may be one of the strategies to pave the way for the clinical translation of radiomic biomarkers.

Regardless of the three resampling methods, most of the features showed an $ICC < 0.5$, *i.e.*, a very low number of features with excellent and good agreement was extracted by a single software from CT and MRI (see Table 3). In this case, the percentage of CT-MRI features highly correlated for both the analyzed software was higher considering resampled images rather than the original ones. The features with good ICC, regardless of the resampling method, are the SHAPE features Volume and Surface Area for both software, plus Maximum3DDiameter SHAPE for LIFEx. Moreover, for both PyRadiomics and LIFEx, the only feature exhibiting excellent ICC is Volume (SHAPE), extracted from external resampled images. This may stem from the use of the same algorithm for image resampling, emphasizing the relevance of harmonizing the preprocessing steps. Considering texture features, we found that NGTDM features Busyness and Strength showed good ICC for both software, with internal and external resampling. In addition to the aforementioned features, texture features with good ICC for the internal and external resampling are RLNU (GLRLN) for LIFEx and ZP (GLZLM) for PyRadiomics. CT-MRI features with good or excellent ICC for PyRadiomics and LIFEx are shown in Table 4.

These results allow us to draw several considerations. As mentioned before, CT and MRI are two different techniques. Hence, our initial hypothesis was that most information embedded within the images would not be directly translatable from one image technique to the other. This hypothesis was corroborated by the analysis results. Vuong et al. had previously shown that SHAPE and texture features are highly correlated between PET/CT and PET/MRI [13]. Moving from nuclear medicine to diagnostic imaging, the same result was not yet known. Even though the correlation between certain CT and MRI SHAPE features could have been expected, as SHAPE features describe tumor morphological aspects, the correlation between CT and MRI

texture features was not taken for granted. It is remarkable especially that two NGTDM features (Busyness and Strength) presented good ICC for both internal and external resampling and this was consistent across both radiomic software.

This study has limitations. First, the correlation we found is limited to the pathological process considered and the kind of images we compared, *i.e.*, NSCLC and contrast-enhanced CT and T1-weighted MRI. Second, while a diversity of CT equipment was used, only one 1.5-T MRI unit with a specific pulse sequence was employed. Third, this is a single-center study with a relatively small sample size. Nevertheless, it is a reliable proof of concept as first of all, currently, just a few centers have incorporated lung MRI into clinical practice, and secondly, it is relatively uncommon to have not only CT but also MRI acquisition for each patient.

In conclusion, we investigated the agreement between LIFEx and PyRadiomics software for two different imaging techniques, explored the correlation between CT-MRI corresponding features calculated with such software, and assessed how these relationships are affected by the resampling method. PyRadiomics and LIFEx are highly in agreement with each other for both MRI and CT (on average 90% of features showed $ICC \geq 0.75$) and approximately 10% of MRI-CT related features (shape and texture) obtained from resampled images exhibited $ICC \geq 0.75$. The impact of the resampling method was clear in the previous points. For both imaging modalities, we observed a decrease in agreement between the two radiomic tools when using their internal resampling methods.

To validate our results, further investigations using a wider multi-center cohort are necessary. These additional studies are also needed to confirm the "identity" of stable features in cross-modality and their generalizability as useful biomarkers. Furthermore, it is essential to single out sequences that are most suitable for radiomics with the aim of establishing a standardized acquisition protocol, especially for MRI images. A wider exploration of the potential of MRI radiomics in lung cancer patients through the use of other sequences (*e.g.*, unenhanced T1-weighted, T2-weighted, short- τ , diffusion metrics) could allow strengthening and/or expanding the identification of the most relevant features. Lastly, it would be interesting to evaluate the stability of our results across different segmentation methods.

As a final remark, in both MRI and CT cases, the study of physical principles and biomedical mechanisms underlying the radiomic features definition (a very challenging issue) and the extension of the presented approach to more modalities, including PET technique [14, 40], will possibly provide more robust biomarkers within a broader multimodality approach.

Abbreviations

CT	Computed tomography
GLCM	Gray level co-occurrence matrix
GLRLM	Gray-level run length matrix
GLZLM	Gray-level size zone matrix
IBSI	Image Biomarker Standardisation Initiative
ICC	Intraclass correlation coefficient
MRI	Magnetic resonance imaging
NGTDM	Neighboring gray-tone difference matrix
NSCLC	Non-small cell lung cancer
PET	Positron emission tomography

Acknowledgements

FB, RFC, ART, MM, IP, and AL thank the INFN-CSNS research project *next AIM* (Artificial Intelligence in Medicine: next steps), <https://www.pi.infn.it/aim/>. Programma di ricerca CNO000013 "National Centre for HPC, Big Data and Quantum Computing" is acknowledged. MIUR Dipartimenti di Eccellenza 2018–2022 project F11118000680001 is also gratefully acknowledged for funding the present research.

Authors' contributions

Conceptualization, AP, GM, CB, OMB, and LP; methodology, AP, GM, FB, RFC, ART, and IP; software, FB, RFC, ART, and IP; validation, FB, RFC, ART, and IP; formal analysis, FB, RFC, ART, and IP; investigation, AP, GM, FB, and RFC; resources, CB, OMB, AL, and LP; data curation, AP, GM, OMB, and CB; writing—original draft preparation, AP, GM, FB, RFC, ART, and IP; writing—review and editing, GMS, GG, ARF, MM, AL, and LP; visualization, CB, OMB, SF, MM, AL, and LP; supervision, CB, OMB, SF, MM, AL, and LP; project administration, SF, AL, and LP; funding acquisition, SF, AL, and LP. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by an institutional grant from Fondazione IRCCS Policlinico San Matteo, funding number: 20190049910, and by the Istituto Nazionale di Fisica Nucleare—INFN*next AIM* project.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to privacy restrictions but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the local Medical Research Ethics Committee of Fondazione IRCCS Policlinico San Matteo (Protocol code P 20130113422), and informed consent was obtained from all participants. Informed consent was obtained from all subjects involved in the study.

Consent for publication

Not applicable.

Competing interests

Andrea Riccardo Filippi discloses speakers' bureau support from Astra Zeneca, MSD Italia (<https://www.msd-italia.it/>), Roche, and Ipsen; an advisory role for Astra Zeneca and Roche; research funding from Astra Zeneca; participation (no financial interest) in sponsored research for Astra Zeneca, Roche, and MSD. The remaining authors declare that they have no competing interests.

Author details

¹Radiology Institute, Fondazione IRCCS Policlinico San Matteo, Pavia 27100, Italy. ²Diagnostic Imaging and Radiotherapy Unit, Department of Clinical, Surgical, Diagnostic, and Pediatric Sciences, University of Pavia, Pavia 27100, Italy. ³Department of Physics, University of Pavia, Via Bassi 6, Pavia 27100, Italy. ⁴Istituto Nazionale Di Fisica Nucleare, Sezione Di Pavia, Pavia 27100, Italy. ⁵Department of Mathematics, University of Pavia, Via Ferrata 5, Pavia 27100, Italy. ⁶Department of Medical Sciences and Infective Diseases, Unit of Respiratory Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia 27100, Italy. ⁷Department of Internal Medicine and Medical Therapeutics, University of Pavia, Pavia 27100, Italy. ⁸Department of Political and Social Sciences, University of Pavia, Pavia 27100, Italy. ⁹Department of Radiation Oncology, Fondazione IRCCS Policlinico San Matteo, Pavia 27100, Italy.

Received: 31 January 2024 Accepted: 10 April 2024

Published online: 17 June 2024

References

- Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>
- Shur J, Blackledge M, D'Arcy J et al (2021) MRI texture feature repeatability and image acquisition factor robustness, a phantom study and in silico study. *Eur Radiol Exp* 5:1–11. <https://doi.org/10.1186/s41747-020-00199-6>
- Yip SS, Aerts HJ (2016) Applications and limitations of radiomics. *Phys Med Biol* 61:R150. <https://doi.org/10.1088/0031-9155/61/13/R150>
- Lacroix M, Froin F, Dirand AS et al (2020) Correction for magnetic field inhomogeneities and normalization of voxel values are needed to better reveal the potential of MR radiomic features in lung cancer. *Front Oncol* 10:43. <https://doi.org/10.3389/fonc.2020.00043>
- Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. *Magn Reson Imaging* 30:1234–1248. <https://doi.org/10.1016/j.mri.2012.06.010>
- Parmar C, Grossmann P, Rietveld D, Rietbergen M, Lambin P, Aerts H (2015) Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol* 5:272. <https://doi.org/10.3389/fonc.2015.00272>
- Lee SH, Cho Hh, Kwon J, Lee HY, Park H (2021) Are radiomics features universally applicable to different organs? *Cancer Imaging* 21:1–10. <https://doi.org/10.1186/s40644-021-00400-y>
- Rizzo S, Botta F, Raimondi S et al (2018) Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2:1–8. <https://doi.org/10.1186/s41747-018-0068-z>
- Sardanelli F (2017) Trends in radiology and experimental research. *Eur Radiol Exp* 1:1–7. <https://doi.org/10.1186/s41747-017-0006-5>
- Mahon RN, Hugo GD, Weiss E (2019) Repeatability of texture features derived from magnetic resonance and computed tomography imaging and use in predictive models for non-small cell lung cancer outcome. *Phys Med Biol* 64:145007. <https://doi.org/10.1088/1361-6560/ab18d3>
- Liberini V, Laudicella R, Balma M et al (2022) Radiomics and artificial intelligence in prostate cancer: new tools for molecular hybrid imaging and theragnostics. *Eur Radiol Exp* 6:27. <https://doi.org/10.1186/s41747-022-00282-0>
- Mayerhoefer ME, Materka A, Langs G et al (2020) Introduction to radiomics. *J Nucl Med* 61:488–495. <https://doi.org/10.2967/jnumed.118.222893>
- Vuong D, Tanadini-Lang S, Huellner MW et al (2019) Interchangeability of radiomic features between [18F]-FDG PET/CT and [18F]-FDG PET/MR. *Med Phys* 46:1677–1685. <https://doi.org/10.1002/mp.13422>
- Tang X, Liang J, Xiang B et al (2022) Positron emission tomography/magnetic resonance imaging radiomics in predicting lung adenocarcinoma and squamous cell carcinoma. *Front Oncol* 12:13. <https://doi.org/10.3389/fonc.2022.803824>
- Thawani R, McLane M, Beig N et al (2018) Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer* 115:34–41. <https://doi.org/10.1016/j.lungcan.2017.10.015>
- Park H, Sholl LM, Hatabu H, Awad MM, Nishino M (2019) Imaging of precision therapy for lung cancer: current state of the art. *Radiology* 293:15–29. <https://doi.org/10.1148/radiol.2019190173>
- Hochegger B, Zanon M, Altmayer S et al (2018) Advances in imaging and automated quantification of malignant pulmonary diseases: a state-of-the-art review. *Lung* 196:633–642. <https://doi.org/10.1007/s00408-018-0156-0>
- Cabini RF, Brero F, Lancia A et al (2022) Preliminary report on harmonization of features extraction process using the ComBat tool in the multi-center "Blue Sky Radiomics" study on stage III unresectable NSCLC. *Insights Imaging* 13:38. <https://doi.org/10.1186/s13244-022-01171-1>
- Rinaldi L, De Angelis SP, Raimondi S et al (2022) Reproducibility of radiomic features in CT images of NSCLC patients: an integrative analysis on the impact of acquisition and reconstruction parameters. *Eur Radiol Exp* 6:2. <https://doi.org/10.1186/s41747-021-00258-6>
- Liu H, Chen R, Tong C, Liang XW (2021) MRI versus CT for the detection of pulmonary nodules: a meta-analysis. *Medicine (Baltimore)* 100:e27270. <https://doi.org/10.1097/MD.00000000000027270>

21. Kauczor HU, Wielpütz MO (2018) MRI of the lung. Springer International Publishing Vol. 6. <https://doi.org/10.1007/978-3-319-42617-4>
22. Sodhi KS, Ciet P, Vasanawala S, Biederer J (2022) Practical protocol for lung magnetic resonance imaging and common clinical indications. *Pediatr Radiol* 52:295–311. <https://doi.org/10.1007/s00247-021-05090-z>
23. Sim AJ, Kaza E, Singer L, Rosenberg SA (2020) A review of the role of MRI in diagnosis and treatment of early stage lung cancer. *Clin Transl Radiat Oncol* 24:16–22. <https://doi.org/10.1016/j.ctro.2020.06.002>
24. Yoon SH, Park CM, Park SJ, Yoon JH, Hahn S, Goo JM (2016) Tumor heterogeneity in lung cancer: assessment with dynamic contrast-enhanced MR imaging. *Radiology* 280:940–948. <https://doi.org/10.1148/radiol.2016151367>
25. Nioche C, Orlhac F, Boughdad S et al (2018) LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res* 78:4786–4789. <https://doi.org/10.1158/0008-5472.CAN-18-0125>
26. Van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
27. Fornaçon-Wood I, Mistry H, Ackermann CJ et al (2020) Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol* 30:6241–6250. <https://doi.org/10.1007/s00330-020-06957-9>
28. Paquier Z, Chao SL, Acquisto A et al (2022) Radiomics software comparison using digital phantom and patient data: IBSI-compliance does not guarantee concordance of feature values. *Biomed Phys Eng Express* 8:065008. <https://doi.org/10.1088/2057-1976/ac8e6f>
29. Bleker J, Roest C, Yakar D, Huisman H, Kwee TC (2023) The effect of image resampling on the performance of radiomics-based artificial intelligence in multicenter prostate MRI. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.28935>
30. Wichtmann BD, Harder FN, Weiss K et al (2023) Influence of image processing on radiomic features from magnetic resonance imaging. *Invest Radiol* 58:199–208. <https://doi.org/10.1097/RLI.0000000000000921>
31. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295:328–338. <https://doi.org/10.1148/radiol.2020191145>
32. Shafiq-ul-Hassan M, Zhang GG, Latifi K et al (2017) Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys* 44:1050–1062. <https://doi.org/10.1002/mp.12123>
33. Shafiq-ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E (2018) Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep* 8:10545. <https://doi.org/10.1038/s41598-018-28895-9>
34. Scalco E, Belfatto A, Mastropietro A et al (2020) T2w-MRI signal normalization affects radiomics features reproducibility. *Med Phys* 47:1680–1691. <https://doi.org/10.1002/mp.14038>
35. Lambin P, Leijenaar RT, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nature Rev Clin Oncol* 14:749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
36. Hoebe KV, Patel JB, Beers AL et al (2020) Radiomics repeatability pitfalls in a scan-rescan MRI study of glioblastoma. *Radiology Artif Intell* 3:e190199. <https://doi.org/10.1148/ryai.2020190199>
37. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R (2018) SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *J Digit Imaging* 31:290–303. <https://doi.org/10.1007/s10278-017-0037-8>
38. Beare R, Lowekamp B, Yaniv Z (2018) Image segmentation, registration and characterization in R with SimpleITK. *J Stat Softw* 86:8. <https://doi.org/10.18637/jss.v086.i08>
39. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
40. Vogl WD, Pinker K, Helbich TH et al (2019) Automatic segmentation and classification of breast lesions through identification of informative multiparametric PET/MRI features. *Eur Radiol Exp* 3:1–13. <https://doi.org/10.1186/s41747-019-0096-3>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.