NARRATIVE REVIEW   **Open Access**

Radiomics: the facts and the challenges of image analysis

Stefania Rizzo[1*] , Francesca Botta[2], Sara Raimondi[3], Daniela Origgi[2], Cristiana Fanciullo[4],
Alessio Giuseppe Morganti[5] and Massimo Bellomi[6]

Abstract

Radiomics is an emerging translational field of research aiming to extract mineable high-dimensional data from clinical images. The radiomic process can be divided into distinct steps with definable inputs and outputs, such as image acquisition and reconstruction, image segmentation, features extraction and qualification, analysis, and model building. Each step needs careful evaluation for the construction of robust and reliable models to be transferred into clinical practice for the purposes of prognosis, non-invasive disease tracking, and evaluation of disease response to treatment. After the definition of *texture* parameters (shape features; first-, second-, and higher-order features), we briefly discuss the origin of the term *radiomics* and the methods for selecting the parameters useful for a radiomic approach, including cluster analysis, principal component analysis, random forest, neural network, linear/logistic regression, and other. Reproducibility and clinical value of parameters should be firstly tested with *internal cross-validation* and then validated on *independent external cohorts*. This article summarises the major issues regarding this multi-step process, focussing in particular on challenges of the extraction of radiomic features from data sets provided by computed tomography, positron emission tomography, and magnetic resonance imaging.

Keywords: Clinical decision-making, Biomarkers, Image processing (computer-assisted), Radiomics, Texture analysis

Key points

- Radiomics is a complex multi-step process aiding clinical decision-making and outcome prediction
- Manual, automatic, and semi-automatic segmentation is challenging because of reproducibility issues
- Quantitative features are mathematically extracted by software, with different complexity levels
- Reproducibility and clinical value of radiomic features should be firstly tested with internal cross-validation and then validated on independent external cohorts

Background

In the new era of *precision medicine*, *radiomics* is an emerging translational field of research aiming to find associations between qualitative and quantitative information extracted from clinical images and clinical data, with or without associated gene expression to support evidence-based clinical decision-making [1]. The concept underlying the process is that both morphological and functional clinical images contain qualitative and quantitative information, which may reflect the underlying pathophysiology of a tissue. Radiomics' analyses can be performed in tumour regions, metastatic lesions, as well as in normal tissues [2].

The radiomics quantitative features can be calculated by dedicated software, which accepts the medical images as an input. Despite many tools developed for this specific task being user-friendly in terms of use, and well performing in terms of calculation time, it is still challenging to carefully check the quality of the input data and to select the optimal parameters to guarantee a reliable and robust output.

The quality of features extracted, their association with clinical data, and also the model derived from them, can be affected by the type of image acquisition, postprocessing, and segmentation.

This article summarises the major issues regarding this multi-step process, focussing in particular on the challenges

* Correspondence: stefania.rizzo@ieo.it
[1]Department of Radiology, IEO, European Institute of Oncology, IRCCS, Milan, IT, Italy
Full list of author information is available at the end of the article

Rizzo *et al. European Radiology Experimental* (2018) 2:36

Page 2 of 8

that the extraction and radiomics' use of imaging features from computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI) generates.

Definition and extraction of image features

Different kind of features can be derived from clinical images. Qualitative semantic features are commonly used in the radiology lexicon to describe lesions [3]. Quantitative features are descriptors extracted from the images by software implementing mathematical algorithms [4]. They exhibit different levels of complexity and express properties firstly of the lesion shape and the voxel intensity histogram, secondarily of the spatial arrangement of the intensity values at voxel level (*texture*). They can be extracted either directly from the images or after applying different filters or transforms (e.g., wavelet transform).

Quantitative features are usually categorised into the following subgroups:

Shape features describe the shape of the traced region of interest (ROI) and its geometric properties such as volume, maximum diameter along different orthogonal directions, maximum surface, tumour compactness, and sphericity. For example, the surface-to-volume ratio of a spiculated tumour will show higher values than that of a round tumour of similar volume.

First-order statistics features describe the distribution of individual voxel values without concern for spatial relationships. These are histogram-based properties reporting the mean, median, maximum, minimum values of the voxel intensities on the image, as well as their skewness (asymmetry), kurtosis (flatness), uniformity, and randomness (entropy).

Second-order statistics features include the so-called *textural features* [5, 6], which are obtained calculating the statistical inter-relationships between neighbouring voxels [7]. They provide a measure of the spatial arrangement of the voxel intensities, and hence of intra-lesion heterogeneity. Such features can be derived from the grey-level co-occurrence matrix (GLCM), quantifying the incidence of voxels with same intensities at a predetermined distance along a fixed direction, or from the *Grey-level run-length matrix* (GLRLM), quantifying consecutive voxels with the same intensity along fixed directions [8].

Higher-order statistics features are obtained by statistical methods after applying filters or mathematical transforms to the images; for example, with the aim of identifying repetitive or non-repetitive patterns, suppressing noise, or highlighting details. These include fractal analysis, Minkowski functionals, wavelet transform, and Laplacian transforms of Gaussian-filtered images, which can extract areas with increasingly coarse texture patterns.

Considering that many parameters can be tuned by the user, hundreds of variables can be generated from a single image.

Most of the abovementioned features are neither original nor innovative descriptors. Indeed, the definition and use of textural features to quantify image properties, as well as the use of filters and mathematical transforms to process signals, date back a few decades [6]. Therefore, the main innovation of radiomics relies on the – omics suffix, originally created for molecular biology disciplines. This refers to the simultaneous use of a large amount of parameters extracted from a single lesion, which are mathematically processed with advanced statistical methods under the hypothesis that an appropriate combination of them, along with clinical data, can express significant tissue properties, useful for diagnosis, prognosis, or treatment in an individual patient (personalisation). Additionally, radiomics takes advantage of the full use of large data-analysis experience developed by other -omics disciplines, as well as by big-data analytics.

Some difficulties arise when the user has to choose which and how many parameters to extract from the images. Each tool calculates a different number of features, belonging to different categories, and the initial choice may appear somehow arbitrary. Nonetheless, methods for data analysis strictly depend on the number of input variables, possibly affecting the final result. One possible approach is to start from all the features provided by the calculation tool, and to perform a preliminary analysis to select the most repeatable and reproducible parameters; to subsequently reduce them by correlation and redundancy analysis [9]. Another approach is to make an *a priori* selection of the features, based on their mathematical definition, focussing on the parameters easily interpretable in terms of visual appearance, or directly connectable to some biological properties of the tissue.

Alternatively, *machine-learning* techniques, underlying the idea that computers may learn from past examples and detect hard-to-discern patterns from large and complex data sets, are emerging as useful tools that may lead to the selection of appropriate features [10–12].

Analysis and model building

Many of the extracted features are redundant. Therefore, initial efforts should focus on identifying appropriate endpoints with a potential clinical application, to select information useful for a specific purpose. Radiomics' analysis usually includes two main steps:
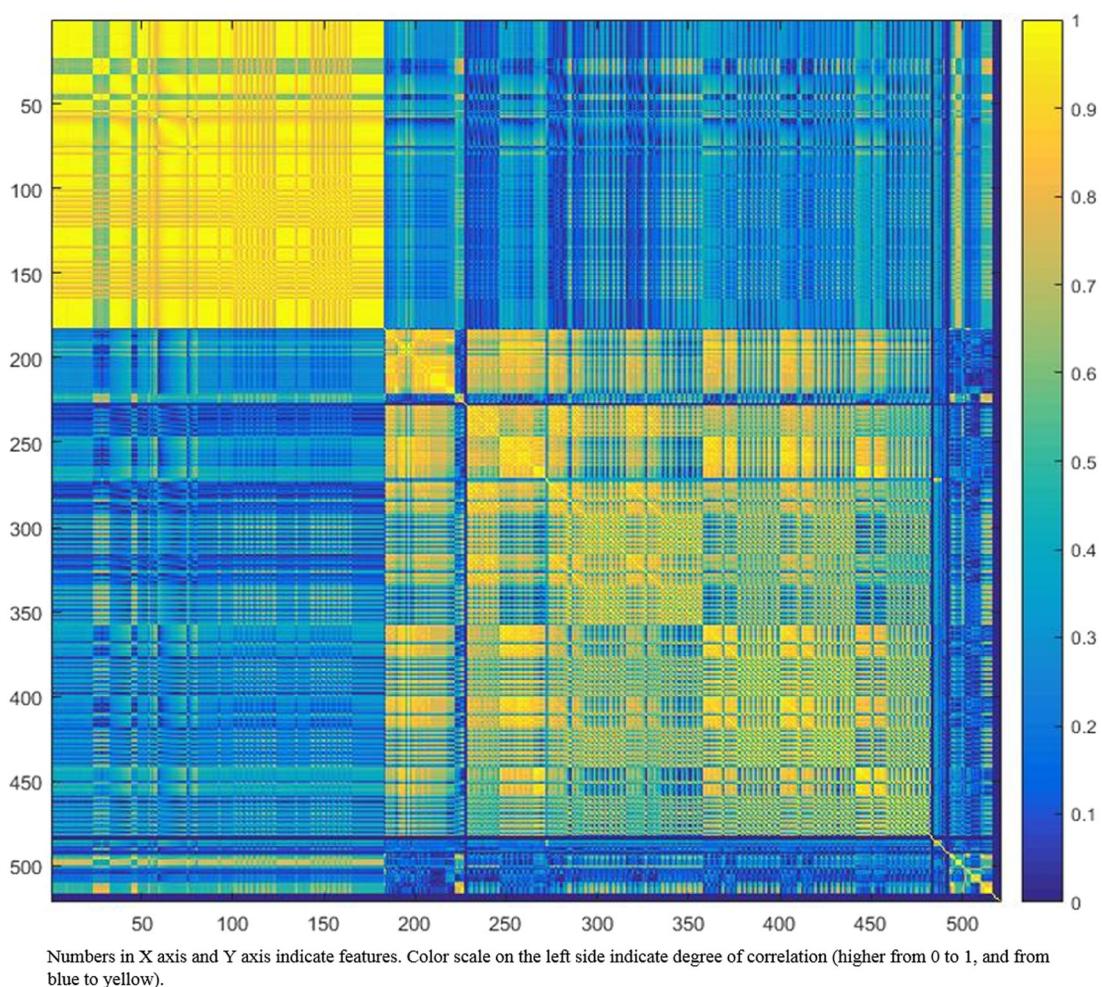
1. Dimensionality reduction and feature selection, usually obtained via unsupervised approaches; and
2. Association analysis with one or more specific outcome(s) via supervised approaches.

Different methods of dimensionality reduction/feature selection and model classification have been compared [13, 14]. The two most commonly used unsupervised approaches are *cluster analysis* [7, 14, 15] and *principal component analysis* (PCA) [13, 16]. Cluster analysis aims to create groups of similar features (clusters) with high intra-cluster redundancy and low intercluster correlation. This type of analysis is usually depicted by a cluster heat map [17], as shown in Fig. 1. A single feature may be selected from each cluster as representative and used in the following association analysis [14, 15]. PCA aims to create a smaller set of maximally uncorrelated variables from a large set of correlated variables, and to explain as much as possible of the total variation in the data set with the fewest possible principal components

[18]. Graphically, the output of PCA consists of score plots, giving an indication for grouping in the data sets for similarity.

All selected features considered reproducible, informative, and non-redundant can then be used for association analysis. According to our experience, an important *caveat* for univariate analysis is multiple testing. The most common way to overcome the multiple testing problem is to use Bonferroni correction or the less conservative false discovery rate corrections [19].

Supervised multivariate analysis consists of building a mathematical model to predict an outcome or response variable. The different analysis approaches depend on the purpose of the study and the outcome category, ranging from statistical methods to data-mining/

Numbers in X axis and Y axis indicate features. Color scale on the left side indicate degree of correlation (higher from 0 to 1, and from blue to yellow).

Fig. 1 Graphic representation of radiomic-feature clustering. This example graph displays the absolute value of the correlation coefficient (ranging from 0 to 1, on the right side, indicating increasing degree of correlation) between each pair of radiomic features (shown as numbers on the two axes). The heat map gives a good visual representation of the high correlation observed for most radiomic features that may be grouped in the same cluster to avoid redundancy. The yellow blocks along the diagonal graphically identify the clusters including highly correlated radiomic features. Blue blocks outside the diagonal visualise the low correlation observed between radiomic features belonging to different clusters. In the present example, two major clusters with different information may be identified, with very high redundancy for radiomic features in the first cluster (high homogeneity of the yellow blocks)

machine-learning approaches, such as random forests [14, 20], neural networks [21], linear regression [21], logistic regression [15], least absolute shrinkage and selection operator [22], and Cox proportional hazards regression [23]. Previous studies comparing different model-building approaches found that the random forest classification method had the highest prognostic performance [13, 14].

Unquestionably, the stability and reproducibility of the model must be assessed before applying a predictive model in a clinical setting. Indeed, it is well known that model fitting is optimal in the training set used to build the model, while validation in an external cohort provides more reliable fitting estimates [24]. The first step in model validation is *internal cross-validation*. However, the best way to assess the potential clinical value of a model is validation with prospectively collected independent cohorts, ideally within clinical trials. This introduces the issue of data sharing among different institutions, creating the need for shared databases to be used as validation sets. To help solve this issue, there are large, publicly available databases, such as The Cancer Genome Atlas (TCGA), including comprehensive multidimensional genomic data and clinical annotations of more than 30 types
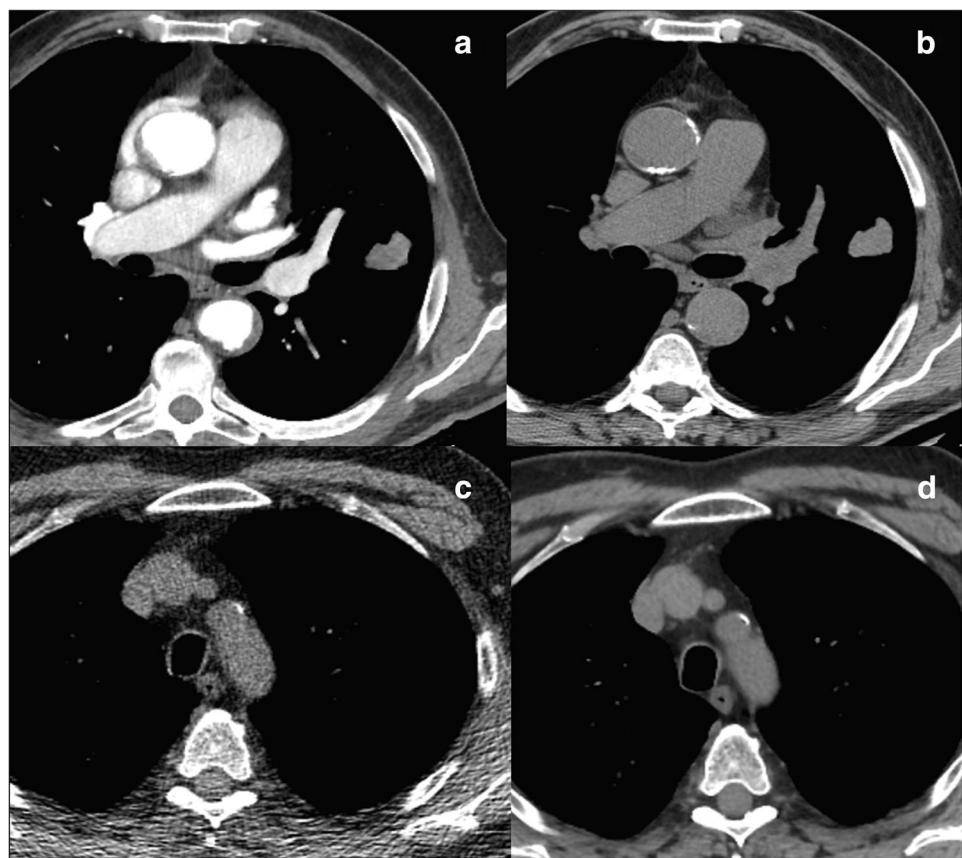
of cancer [25]. Likewise, the Cancer Imaging Archive is a publicly available resource hosting the imaging data of patients in the TCGA database. These images can be used as valuable sources for both hypothesis generating and validation purposes [26].

Notably, patient parameters may influence image features via a direct causal association or exert a confounding effect on statistical associations. For instance, smoking-related lung cancers differ from lung cancers in non-smokers [27].

Moreover, since models need validation to be preferably performed on external and independent groups of patients, the comparability of features extracted from images with different parameters and segmented with different techniques is challenging and may affect the final performance of the model itself.

Impact of image acquisition and reconstruction

Routine clinical imaging techniques show a wide variation in acquisition parameters, such as: image spatial resolution; administration of contrast agents; kVp and mAs (among others) for CT (Fig. 2); type of sequence, echo time, repetition time, number of excitations and many other sequence parameters for MRI. Furthermore,

Fig. 2 Axial computed tomography images showing differences in the same acquisition plane between a contrast-enhanced (**a**) and an unenhanced image (**b**), as well as for different radiation doses, lower in (**c**), and higher in (**d**)

different vendors offer different reconstruction algorithms, and reconstruction parameters are customised at each institution, with possible variations in individual patients. All these variables affect image noise and texture, and consequently the value of the radiomic features. As a result, features obtained from images acquired at a single institution using different acquisition protocols, or acquired at different institutions with different scanners in different patient populations, may be affected by different parameters, rather than reflecting different biological properties of tissues. Finally, some acquisition and reconstruction settings may yield to unstable features, thus showing different values when extracted from repeated measurements under identical conditions.

An approach to overcome this limitation may be to exclude from the beginning the features highly influenced by the acquisition and reconstruction parameters. This can be achieved by integrating information from the literature and from dedicated experimental measurements, taking into account the peculiarity of each imaging modality.

CT
Standard CT phantoms, like those proposed by the American Association of Physicists in Medicine [28], allow the evaluation of imaging performance and the assessment of how far image quality depends on the adopted technique. Despite not being intended for this, they may provide useful information on the parameters potentially affecting image texture. For instance, a decrease in slice thickness reduces the photon statistics within a slice (unless mAs or kVp are increased accordingly), thereby increasing image noise. The axial field of view and reconstruction matrix size determine the pixel size and hence the spatial sampling in the axial plane, which has an impact on the description of heterogeneity. The reduction of pixel size increases image noise (when the other parameters are kept unchanged), but increases spatial resolution.

When considering spiral CT acquisition, pitch is a variable that influences image noise, making difficult the comparison between different scanners and vendors. Thus, non-spiral (axial) acquisitions are necessary for these comparisons. Likewise, clinical conditions, such as the presence of artifacts due to metallic prostheses, may affect image quality and impair quantitative analysis [29]. Furthermore, electronic density quantification expressed as Hounsfield Units may vary with the reconstruction algorithm [30] or scanner calibration.

Thus, to study in detail the effects of acquisition settings and reconstruction algorithms on radiomic features, more sophisticated phantoms are required. For example, the Credence Cartridge Radiomics phantom, including different cartridges, each of them exhibiting a

different texture, was developed to test inter-scanner, intra-scanner, and multicentre variability [31], as well as the effect of different acquisition and reconstruction settings on feature robustness [4]. Another possibility is to develop customised phantoms [32] resembling the anatomic districts of interest, embedding inserts simulating tissues with different texture and size, and located at different positions, to test protocols under real clinical conditions.

Alternatively, many authors have investigated features of robustness and stability on clinical images by undertaking test-retest studies [33], or comparing the results obtained with different imaging settings and processing algorithms [34]. These studies conclude that there is still the need for dedicated investigations to select features with sufficient dynamic range among patients, with intra-patient reproducibility and low sensitivity to image acquisition and reconstruction protocols [15].
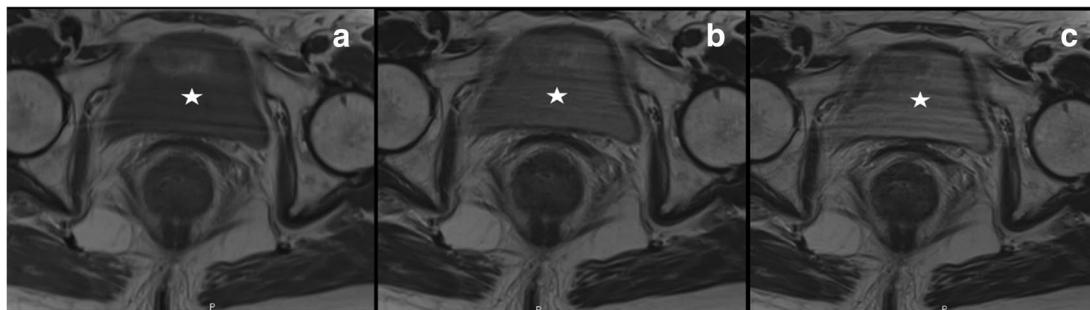
PET
Texture analysis on PET images poses additional challenges. PET spatial resolution is in general worse than that of CT, because of low accuracy in describing the spatial distribution of VI, which radiomic features aim to quantify. This relies on different physical phenomena, different technologies used for radiation detection, and patient motion. Less accurate data may fail in generating significant association with biological and clinical endpoints, or may require an increased number of patients.

Of note, the VI, expressed in terms of standardised uptake value (SUV) can be scanner dependent. For example, modelling or not the detector response in the reconstruction algorithm leads to a lymph node SUV_{mean} difference of 28% [35]. Furthermore, for the same scanner model, SUV differences (hence radiomic-feature differences) may be due to acquisition at different times post injection, patient blood glucose level and presence of inflammation [36].

Previous studies provided data to select the most appropriate procedures and radiomic PET features [37–39]. For example, voxel size was shown to be the most important source of variability for a large number of features, whereas the *entropy* feature calculated from the GLCM was robust with respect to acquisition and reconstruction parameters, post-filtering level, iteration number, and matrix size [35].

For dedicated experimental measurements, phantoms routinely used for PET scanner quality control may be used. For instance, the NEMA Image Quality phantom has been used to assess the impact of noise on textural features when varying reconstruction settings [37, 40], whereas homogeneous phantoms have been used to test stability [41]. To our knowledge, commercial phantoms customised for testing radiomic-feature performance in the presence of inhomogeneous activity distributions are

Fig. 3 Axial T2-weighted images of the pelvis, acquired keeping unchanged all the parameters, with only exception of the echo time, which was 34 ms in (**a**), 90 ms in (**b**), and 134 ms in (**c**), showing that even one single parameter can change the signal intensity of tissues and fluids, as clearly depicted by the signal of the bladder (white star), with higher and higher signal intensity from **a** to **b** to **c**

not yet available, but home-made solutions have been described [41].

Scanner calibration and protocol standardisation are necessary to allow for multicentre studies and model generalisability [9, 42]. Harmonisation methods are emerging to allow gathering and comparing data from different centres, although they are not yet largely applied in clinical studies [35].

MRI
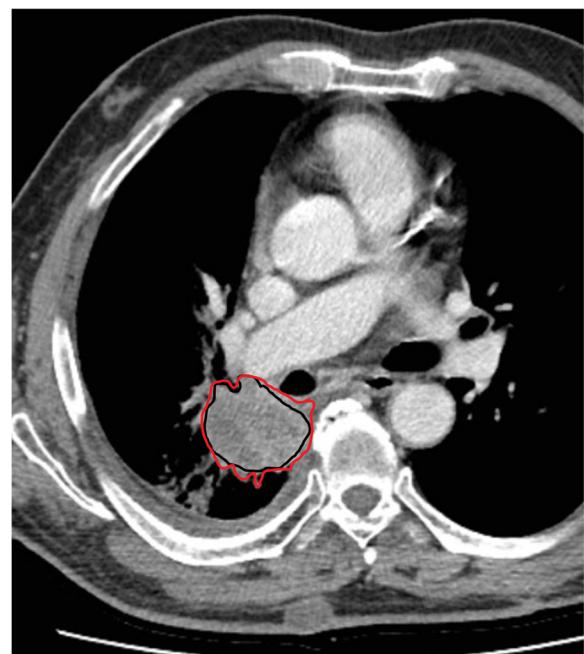
The signal intensity in MRI arises from a complex interaction of intrinsic tissue properties, such as relaxation times as well as multiple parameters related to scanner properties, acquisition settings, and image processing. For a given T1- or T2-weighted sequence, voxel intensity does not have a fixed tissue-specific numeric value. Even when scanning the same patient in the same position with the same scanner using the same sequence in two or more sessions, signal intensity may change (Fig. 3), whereas tissue contrast remains unaltered [43].

Without a correction for this effect, a comparison of radiomic features among patients may lose significance as it depends on the numeric value of voxel intensity. One possibility is to focus texture analysis on radiomic features quantifying the relationship between voxel intensities, where numerical values do not depend on the individual voxel intensity; another is to make a compensation (normalisation) before performing quantitative image analysis [43].

Current studies investigating the impact of MRI acquisition parameters on radiomic-feature robustness address the complexity of the technique and the low availability of proper phantoms. The available data suggest that texture features are sensitive to variations of acquisition parameters: the higher the spatial resolution, the higher the sensitivity [44]. A trial assessing radiomic features obtained on different scanners at different institutions or with different parameters concluded that comparisons should be treated with care [45].

Impact of image segmentation

Segmentation is a critical step of the radiomics process because data are extracted from the segmented volumes. This is challenging because many tumours show unclear borders. It is contentious because there is no consensus on the need to seek either the ground truth or reproducibility of segmentation [1]. Indeed, many authors consider manual segmentation by expert readers the ground truth despite high inter-reader variability. This method is also labour intensive (Fig. 4) and not always feasible for radiomics' analysis, requiring very large data sets [46].

Fig. 4 An example of manual segmentation of lung cancer on computed tomography images. Although manual segmentation is often considered ground truth, this image shows red and black regions of interest delineated by two different readers for the same tumour

Automatic and semi-automatic segmentation methods have been developed across imaging modalities and different anatomical regions. Common requirements include maximum automaticity with minimum operator interaction, time efficiency, accuracy, and boundary reproducibility. Some segmentation algorithms rely on region-growing methods that require an operator to select a seed point within the volume of interest [47]. These methods work well for relatively homogeneous lesions, but show the need for intensive user correction for inhomogeneous lesions. For example, most stage I and stage II lung tumours present as homogenous, high-intensity lesions on a background of low-intensity lung parenchyma [48, 49] and, therefore, can be automatically segmented with high reproducibility and accuracy. However, for partially solid, ground-glass opacities, nodules attached to vessels and to the pleural surface, automatic segmentation is burdened by low reproducibility [50].

Other segmentation algorithms include *level-set methods* that represent a contour as the zero-level set of a higher dimensional function (level-set function), then the method formulates the motion of the contour as the evolution of the level-set function [51]. *Graph-cut methods* construct an image-based graph and accomplish a globally optimal solution of energy minimisation functions, but they are computationally expensive [52] and may lead to over-segmentation [53]. *Active contour (snake) algorithms* work like a stretched elastic band. The starting points are drawn around the lesion; then move through an iterative process to a point with the lowest energy function value. These algorithms may lead the snake to undesired locations because they depend on an optimal starting point and are sensitive to noise [54]. Semi-automatic segmentation algorithms do a graph search through local active contour analysis, while their cost function is minimised using dynamic programming. Nonetheless, the semi-automaticity still requires human interaction [55].

As shown, there is still no universal segmentation algorithm for all image applications, and new algorithms are under evaluation to overcome these limitations [56–58]. Indeed, some features may show stability and reproducibility using one segmentation method, but not another.

Conclusions

To summarise, staying in the present while looking into the future, on the one hand, investigators should put efforts in careful selection of robust features for their own models; on the other hand, the scientific community should put efforts towards standardisation, keeping in mind that appropriate statistical approaches will minimise spurious relationships and lead to more accurate and reproducible results.

These will be unavoidable steps towards the construction of generalisable prognostic and predictive models that will effectively contribute to clinical decision-making and treatment management.

Author details
[1]Department of Radiology, IEO, European Institute of Oncology, IRCCS, Milan, IT, Italy. [2]Medical Physics, European Institute of Oncology, Milan, Italy. [3]Division of Epidemiology and Biostatistics, European Institute of Oncology, Milan, Italy. [4]Università degli Studi di Milano, Postgraduate School in Radiodiagnostics, Milan, Italy. [5]Radiation Oncology Center, School of Medicine, Department of Experimental, Diagnostic and Specialty Medicine – DIMES, University of Bologna, Bologna, Italy. [6]Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, Milan, Italy.

References
1. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
2. Lambin P, Leijenaar RTH, Deist T et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762
3. Rizzo S, Petrella F, Buscarino V et al (2016) CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. Eur Radiol 26:32–42
4. Larue RTHM, van Timmeren JE, de Jong EEC et al (2017) Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol 56:1544–1553

5.   Ergen B, Baykara M (2014) Texture based feature extraction methods for content based medical image retrieval systems. Biomed Mater Eng 24:3055–3062.
6.   Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3:610–621
7.   Balagurunathan Y, Kumar V, Gu Y et al (2014) Test-retest reproducibility analysis of lung CT image features. J Digit Imaging 27:805–823
8.   Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4:172–179
9.   Ollers M, Bosmans G, van Baardwijk A et al (2008) The integration of PET–CT scans from different hospitals into radiotherapy treatment planning. Radiother Oncol 87:142–146
10. Suzuki K (2017) Overview of deep learning in medical imaging. Radiol Phys Technol 10:257–273
11. Peeken JC, Bernhofer M, Wiestler B et al (2018) Radiomics in radiooncology—challenging the medical physicist. Phys Med 48:27–36
12. Giger ML (2018) Machine learning in medical imaging. J Am Coll Radiol 15:512–520
13. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F (2017) Radiomics-based prognosis analysis for non-small cell lung cancer. Sci Rep 7:46349
14. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ (2015) Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087
15. Rizzo S, Botta F, Raimondi S et al (2018) Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months. Eur Radiol.
16. Huynh E, Coroller TP, Narayan V et al (2017) Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT. PLoS One 12:e0169172
17. Wilkinson L, Friendly M (2009) The history of the cluster heat map. Am Stat 63:179–184
18. Jolliffe IT (2002) Principal component analysis, Series: Springer Series in Statistics, 2nd edn. Springer, New York, p 487
19. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9:811–818
20. Breiman L (2001) Random forests. Mach Learn 45:5–32
21. Eschrich S, Yang I, Bloom G et al (2005) Molecular staging for survival prediction of colorectal cancer patients. J Clin Oncol 23:3526–3535
22. Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol. 58:267–288
23. Shedden K, Taylor JM, Enkemann SA et al (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 14:822–827.
24. Harrell FE Jr, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361–387
25. Lee H, Palm J, Grimes SM, Ji HP (2015) The cancer genome atlas clinical explorer: a web and mobile interface for identifying clinical-genomic driver associations. Genome Med 7:112
26. Clark K, Vendt B, Smith K et al (2013) The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26:1045e57
27. Panth KM, Leijenaar RT, Carvalho S et al (2015) Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. Radiother Oncol 116:462–466
28. McCollough C, Bakalyar DM, Bostani M et al (2014) Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT: the report of AAPM task group 220. AAPM Rep 2014:6–23
29. Dalal T, Kalra MK, Rizzo SM et al (2005) Metallic prosthesis: technique to avoid increase in CT radiation dose with automatic tube current modulation in a phantom and patients. Radiology 236:671–675
30. Rizzo SM, Kalra MK, Schmidt B et al (2005) CT images of abdomen and pelvis: effect of nonlinear three-dimensional optimized reconstruction algorithm on image quality and lesion characteristics. Radiology 237:309–315
31. Mackin D, Fave X, Zhang L et al (2015) Measuring computed tomography scanner variability of radiomics features. Invest Radiol 50:757–765
32. Theodorakou C, Horrocks JA, Marshall NW, Speller RD (2004) A novel method for producing x-ray test objects and phantoms. Phys Med Biol 49:1423–1438
33. van Timmeren JE, Leijenaar RTH, van Elmpt W et al (2016) Test-retest data for radiomic feature stability analysis: generalizable or study-specific? Tomography 2:361–365
34. Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E (2016) Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 279:185–194
35. Reuzé S, Schernberg A, Orlhac F et al (2018) Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls and challenges. Int J Radiat Oncol Biol Phys. https://doi.org/10.1016/j.ijrobp.2018.05.022
36. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2017) Characterization of PET/CT images using texture analysis: the past, the present…any future? Eur J Nucl Med Mol Imaging 44:151–165
37. Shiri I, Rahmin A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A (2017) The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. Eur Radiol 27:4498–4509
38. Altazi BA, Zhang GG, Fernandez DC et al (2017) Reproducibility of F18-FDG PET radiomic features for different cervical tumour segmentation methods, gray-level discretization, and reconstruction algorithm. J Appl Clin Med Phys 18:32–48
39. Reuzè S, Orlhac F, Chargari C et al (2017) Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget 8:43169–43179
40. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE (2015) Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. J Med Imaging (Bellingham) 2:041002. https://doi.org/10.1117/1.JMI.2.4.041002
41. Forgacs A, Pall Jonsson H, Dahlbom M et al (2016) A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images. PLoS One 11:e0164113
42. Boellaard R (2009) Standards for PET image acquisition and quantitative data analysis. J Nucl Med 50:11S–20S
43. Madabhushi A, Udupa JK (2006) New methods of MR image intensity standardization via generalized scale. Med Phys 33:3426–3434
44. Mayerhoefer M, Szomolanyi P, Jirak D, Materka A, Trattnig S (2009) Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. Med Phys 36:1236–1243
45. Lerski RA, Schad LR, Luypaert R et al (1999) Multicentre magnetic resonance texture analysis trial using reticulated foam test objects. Magn Reson Imaging 17:1025–1031
46. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248
47. Hojjatoleslami S, Kittler J (1998) Region growing: a new approach. IEEE Trans Image Process 7:1079–1084
48. Kalef-Ezra J, Karantanas A, Tsekeris P (1999) CT measurement of lung density. Acta Radiol 40:333–337
49. Sofka M, Wetzl J, Birkbeck N et al (2011) Multi-stage learning for robust lung segmentation in challenging CT volumes. Med Image Comput Comput Assist Interv 14:667–674
50. Knollmann FD, Kumthekar R, Fetzer D, Socinski MA (2014) Assessing response to treatment in non-small-cell lung cancer: role of tumor volume evaluated by computed tomography. Clin Lung Cancer 15:103–109
51. Gao H, Chae O (2010) Individual tooth segmentation from CT images using level set method with shape and intensity prior. Pattern Recognit 43:2406–2417
52. Chen X, Udupa JK, Bagci U, Zhuge Y, Yao J (2012) Medical image segmentation by combining graph cuts and oriented active appearance models. IEEE Trans Image Process 21:2035–2046.
53. Ye X, Beddoe G, Slabaugh G (2010) Automatic graph cut segmentation of lesions in CT using mean shift superpixels. Int J Biomed Imaging 2010:983963. https://doi.org/10.1155/2010/983963
54. Suzuki K, Kohlbrenner R, Epstein ML, Obajuluwa AM, Xu J, Hori M (2010) Computer-aided measurement of liver volumes in CT by means of geodesic active contour segmentation coupled with level-set algorithms. Med Phys 37:2159
55. Lu K, Higgins WE (2007) Interactive segmentation based on the live wire for 3D CT chest image analysis. Int J Comput Assist Radiol Surg 2:151–167
56. Tan Y, Schwartz LH, Zhao B (2013) Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. Med Phys 40:043502
57. Sun S, Bauer C, Beichel R (2012) Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. IEEE Trans Med Imaging 31:449–460
58. Velazquez ER, Parmar C, Jermoumi M et al (2013) Volumetric CT-based segmentation of NSCLC using 3D-slicer. Sci Rep 3:3529