



HAL
open science

Responsible Radiomics Research for Faster Clinical Translation

Martin Vallières, Alex Zwanenburg, Bogdan Badic, Catherine Cheze-Le Rest,
Dimitris Visvikis, Mathieu Hatt

► **To cite this version:**

Martin Vallières, Alex Zwanenburg, Bogdan Badic, Catherine Cheze-Le Rest, Dimitris Visvikis, et al..
Responsible Radiomics Research for Faster Clinical Translation. *Journal of Nuclear Medicine*, 2017,
10.2967/jnumed.117.200501 . hal-01659280

HAL Id: hal-01659280

<https://hal.science/hal-01659280v1>

Submitted on 8 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Responsible Radiomics Research for Faster Clinical Translation

Running title: Responsible Radiomics Research

Martin Vallières¹, Alex Zwanenburg^{2,3}, Bodgan Badic¹, Catherine Cheze-Le Rest¹, Dimitris Visvikis¹,
Mathieu Hatt¹

(1) LaTIM, INSERM, UMR 1101, IBSAM, UBO, UBL, Brest, France.

(2) National Center for Tumor Diseases (NCT), Dresden, Germany.

(3) German Cancer Research Center (DKFZ), Heidelberg, Germany.

Corresponding author:

Mathieu Hatt

IBRBS, Faculté de Médecine

22 rue Camille Desmoulins

29238 Brest

Email: mathieu.hatt@inserm.fr

Phone: 0033298018111

First author:

Martin Vallières

IBRBS, Faculté de Médecine

22 rue Camille Desmoulins

29238 Brest

Email: martin.vallieres@univ-brest.fr

Phone: 0033298018128

It is now recognized that intratumoral heterogeneity is associated with more aggressive tumor phenotypes, leading to poor patient outcomes (1). Medical imaging plays a central role in related investigations, as radiological images are routinely acquired during cancer management. Imaging modalities such as 2-deoxy-2-18F-fluoro-D-glucose (18F-FDG) positron emission tomography (PET), X-ray computed tomography (CT) and magnetic resonance imaging (MRI) are minimally invasive and would constitute an immense source of potential data for decoding tumor phenotypes (2). Computer-aided diagnosis methods and systems exploiting medical images have been developed for decades, but their wide clinical implementation has been hampered by false-positive rates (3). As a consequence, routine clinical exploitation of images still mostly consists of visual/manual assessments. Nowadays, the development of machine learning techniques and the rise of computational power allow for the exploitation of a very large amount of quantitative features (4). This has led to the development of a new incarnation of computer-aided diagnosis, “radiomics”, which refers to the characterization of tumor phenotypes via the extraction of high-dimensional mineable data – e.g., morphological, intensity-based, fractal-based, textural features, etc. – from medical images and whose subsequent analysis aims at supporting clinical decision-making.

A first proof-of-concept study dedicated to the prediction of tumor outcomes using PET radiomics-based multivariable models built via machine learning was published in 2009 (5). The term “radiomics” was then first employed in 2010 to describe how imaging features can reflect gene expression (6). Other early radiomics studies followed (7,8), including some that highlighted early on that the reliability of existing features is affected by acquisition protocols, reconstruction, test-retest, pre-processing or segmentation (9–13). The overall framework of radiomics was then explicitly described in 2012 (14) and in the years that followed, this new emerging field experienced an exponential growth (15).

In the context of precision oncology, the radiomics workflow for the construction of predictive or prognostic models consists of 3 major steps (Fig. 1A): i) *Medical image acquisition*; ii) *Computation of*

radiomics features; and iii) *Statistical analysis and machine learning*. In order for the models to be applied to new patients for treatment personalization, a prospective model evaluation (preferably in a multi-center set-up) is necessary.

Radiomics research has already shown great promise to support clinical decision-making. However, translation of radiomics-based strategies to routine practice has so far not been carried out. This could be partly attributed to the low reproducibility potential of most of current studies. The workflow for computing features is complex and involves many steps (Fig. 1B), which often leads to incomplete reporting of methodological information (e.g., texture matrix design choices and gray-level discretization method). As a consequence, very few radiomics studies in the current literature can be reproduced from start to end. Other major issues include the limited number of patients available for radiomics research, high false-positive rates (similarly to analogous computer-aided diagnosis methods) and the reporting of over-optimistic results, which affect the generalizability of the conclusions reached in current studies.

Medical imaging journals are currently overwhelmed by a large volume of radiomics-related manuscripts of variable quality and associated clinical value. The aim of this editorial is to present guidelines that we think could improve: i) the reporting quality and therefore the reproducibility of radiomics studies; and ii) the statistical quality of radiomics analyses. These guidelines could serve both authors of such studies but also reviewers having to assess their appropriateness for publication.

Guidelines for improving quality of radiomics analyses

The complexity of the radiomics workflow increases the need of standardizing computation methods (16–19). Since September 2016, about 55 researchers from 19 institutions in 8 countries have participated to the Image Biomarker Standardization Initiative (IBSI), which aims at standardizing both the computation of features and the image processing steps required prior to feature extraction (e.g., image interpolation and discretization). A simple digital phantom with few discrete image intensities

was first used to standardize the computation of 172 features from 11 different categories. Secondly, a set of CT images of a lung cancer patient was used to standardize the image processing steps. The initiative is now reaching completion, and a consensus regarding image processing and computation of features was reached over time (20,21). However, more work is likely necessary to define and benchmark MRI- and PET-specific image processing steps. Nonetheless, the standardized workflow (Fig. 1B) along with benchmark values could serve as a calibration tool for future investigations. Ultimately, it could also lead to standardized software solutions available to the community, as the widespread use of standardized computation methods would greatly enhance the reproducibility potential of radiomics studies. It would also be desirable that existing software update their code to conform to future standards to be established by the IBSI. Furthermore, it is essential to rely on supplementary material (usually allowed in most journals) to provide complete methodological details, including the comprehensive description of image acquisition protocols, sequence of operations, image post-acquisition processing, tumor segmentation, image interpolation, image re-segmentation and discretization, formulas for the calculation of features, and benchmark calibrations. Table 1 provides guidelines on feature computation details to be reported in radiomics studies.

Following feature extraction, statistical analysis relates features to clinical outcomes. No consensus exists about what defines “good” radiomics studies. For example, the demonstration that a newly designed feature is strongly associated with a given outcome, or that a novel radiomics method is shown to hold great potential, may be of interest if compared to the most reproducible and robust features or clinical information already used. Nonetheless, for the construction of prediction models via multivariable analysis, basic requirements include: i) all methodological details and clinical information are clearly reported/described, in order to facilitate reproducibility and comparison with other studies and meta-analyses; and ii) radiomics-based models are tested in sufficiently large patient datasets, distinct from teaching (training and validation) sets to statistically demonstrate their efficacy over conventional models (existing biomarkers, tumor volume, cancer stage, etc.). Ideally, for optimal

reproducibility potential, all data and programming code related to the study is also made available to the community. Table 2 provides guidelines based on the “Radiomics Quality Score” (radiomics.world) that can help in evaluating the quality of radiomics studies. More guidelines regarding reproducible prognostic modeling can also be found in the statement for Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) (22).

Responsible research is the key

Some guiding principles already exist to help radiomics scientists to further implement the responsible research paradigm into their current practice. For one, the Responsible Research and Innovation website (rri-tools.eu) provides useful guidelines. Furthermore, a concise set of principles for better scientific data management and stewardship has been defined, the “FAIR Guiding Principles” (23), stating that all research objects should be: i) *F*indable; ii) *A*ccessible; iii) *I*nteroperable; and iv) *R*eusable. Implementation of the FAIR principles within the radiomics field can facilitate its faster clinical translation. Many research tools and online repositories already implement a variety of aspects of the FAIR principles (23), and we can add two other tools of interest: i) The Cancer Imaging Archive (cancerimagingarchive.net), a service which de-identifies and hosts medical images accessible for public download; and ii) The Radiomics Ontology (bioportal.bioontology.org/ontologies/RO), a repository on the National Center for Biomedical Ontology BioPortal aiming to improve the interoperability of radiomics analyses via consistent tagging of radiomics features, segmentation algorithms and imaging filters. This ontology could provide a standardized way of reporting on radiomics data and methods and would summarize the implementation details of a given radiomics workflow (e.g., Table 1) in a more concise manner.

To conclude, initial pioneer studies in radiomics have paved the way to an exciting field and to most promising methods to better personalize cancer treatments. Yet, better standardization, transparency and sharing practices in the radiomics community are required to improve the quality of

published studies and to achieve a faster clinical translation. *Responsible radiomics research* is the best way to achieve this goal, and it can be summarized into three working principles that we should all try to follow as a research community: i) design and conduct high-quality radiomics research; ii) write and present fully transparent radiomics research; and iii) share data and methods.

References

1. Nowell PC. Tumor progression: a brief historical perspective. *Semin Cancer Biol.* 2002;12:261–266.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278:563–577.
3. Giger ML, Chan H-P, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Med Phys.* 2008;35:5799–5820.
4. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *N Engl J Med.* 2016;375:1216–1219.
5. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42:1162–1171.
6. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol.* 2010;65:517–521.
7. Tixier F, Cheze Le Rest C, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.
8. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol.* 2012;102:239–245.
9. Guggenbuhl P, Chappard D, Garreau M, Bansard J-Y, Chales G, Rolland Y. Reproducibility of CT-based bone texture parameters of cancellous calf bone samples: influence of slice thickness. *Eur J Radiol.* 2008;67:514–520.
10. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. *Med Phys.* 2009;36:1236–1243.
11. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol.* 2010;49:1012–1016.
12. Tixier F, Hatt M, Cheze Le Rest C, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J Nucl Med.* 2012;53:693–700.
13. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ¹⁸F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging.* 2013;40:1662–1671.

14. Lambin P, Velazquez ER, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446.
15. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61:R150-R166.
16. Hatt M, Tixier F, Pierce L, Kinahan PE, Cheze Le Rest C, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging*. 2017;44:151–165.
17. Leijenaar RTH, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep*. 2015;5:11075.
18. 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*. 2015;2:041002.
19. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. *Sci Rep*. 2017;7:358.
20. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv161207003*. 2016.
21. Zwanenburg A. EP-1677: Multicentre initiative for standardisation of image biomarkers. *Radiother Oncol*. 2017;123:S914–S915.
22. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
23. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.

Figures

Figure 1. Radiomics workflow. (A) From medical imaging acquisition to treatment personalization. (B) Workflow of computation of radiomics features. Reprinted from (20). ©2016-2017 IBSI. Creative Commons Attribution 4.0 International License.

Tables

Table 1. Reporting guidelines on the computation of radiomics features.

GENERAL	
Image acquisition	Acquisition protocols and scanner parameters: equipment vendor, reconstruction algorithms and filters, field of view and acquisition matrix dimensions, MRI sequence parameters, PET acquisition time and injected dose, CT x-ray energy (kVp) and exposure (mAs), etc.
Volumetric analysis	Imaging volumes are analyzed as separate images (2D) or as fully-connected volumes (3D).
Workflow structure	Sequence of processing steps leading to the extraction of features.
Software	Software type and version of code used for the computation of features.
IMAGE PRE-PROCESSING	
Conversion	How data were converted from input images: e.g, conversion of PET activity counts to SUV, calculation of ADC maps from raw DW-MRI signal, etc.
Processing	Image processing steps taken after image acquisition: e.g., noise filtering, intensity non-uniformity correction in MRI, partial-volume effect corrections, etc.
ROI SEGMENTATION^{a,b}	
How regions of interests (ROIs) were delineated in the images: software and/or algorithms used, how many different persons and what expertise (specialty, experience), how a consensus was obtained if several persons carried out the segmentation, in automatic or semi-automatic mode, etc.	
INTERPOLATION	
Voxel dimensions	Original and interpolated voxel dimensions.
Image interpolation method	Method used to interpolate voxels values (e.g, linear, cubic, spline, etc.) as well as how original and interpolated grids were aligned.
Intensity rounding	Rounding procedures for non-integer interpolated gray levels (if applicable), e.g., rounding of Hounsfield units in CT imaging following interpolation.
ROI interpolation method	Method used to interpolate ROI masks. Definition of how original and interpolated grids were aligned.
ROI partial volume	Minimum partial volume fraction required to include an interpolated ROI mask voxel in the interpolated ROI (if applicable): e.g., a minimum partial volume fraction of 0.5 when using linear interpolation.
ROI RE-SEGMENTATION	
Inclusion/exclusion criteria	Criteria for inclusion and/or exclusion of voxels from the ROI intensity mask (if applicable), e.g., the exclusion of voxels with Hounsfield units values outside a pre-defined range inside the ROI intensity mask in CT imaging.
IMAGE DISCRETIZATION	
Discretization method	Method used for discretizing image intensities prior to feature extraction: e.g., fixed bin number, fixed bin width, histogram equalization, etc.
Discretization parameters	Parameters used for image discretization: the number of bins, the bin width and minimal value of discretization range, etc.
FEATURE CALCULATION	
Features set	Description and formulas of all calculated features.
Features parameters	Settings used for the calculation of features: voxel connectivity, with or without merging by slice, with or without merging directional texture matrices, etc.
CALIBRATION	
Image processing steps	Specifying which image processing steps match the benchmarks of the IBSI.
Features calculation	Specifying which feature calculations match the benchmarks of the IBSI.

^a In order to reduce inter-observer variability, automatic and semi-automatic methods are favored.

^b In multimodal applications (e.g., PET/CT, PET/MRI, etc.) ROI definition may involve the propagation of contours between modalities via co-registration. In that case, the technical details of the registration should also be provided.

Table 2. Quality factors in radiomics studies.

IMAGING	
Standardized imaging protocols	Imaging acquisition protocols are well described and ideally similar across patients. Alternatively, methodological steps are taken towards standardizing them.
Imaging quality assurance	Methodological steps are taken to only incorporate acquired images of sufficient quality.
Calibration	Computation of radiomics features and image processing steps match the benchmarks of the IBSI.
EXPERIMENTAL SETUP	
Multi-institutional/external datasets	Model construction and/or performance evaluation is carried out using cohorts from different institutions, ideally from different parts of the world.
Registration of prospective study	Prospective studies provide the highest level of evidence supporting the clinical validity and usefulness of radiomics models.
FEATURE SELECTION	
Feature robustness	The robustness of features against segmentation variations and varying imaging settings (e.g., noise fluctuations, inter-scanner differences, etc.) is evaluated. Unreliable features are discarded.
Feature complementarity	The inter-correlation of features is evaluated. Redundant features are discarded.
MODEL ASSESSMENT	
False discovery corrections	Correction for multiple testing comparisons (e.g., Bonferroni or Benjamini-Hochberg) is applied in univariate analysis.
Estimation of model performance	The teaching dataset is separated into training and validation set(s) to estimate optimal model parameters. Example methods include bootstrapping, cross-validation, random sub-sampling, etc.
Independent testing	A testing set distinct from the teaching set is used to evaluate the performance of complete models (i.e., without retraining and without adaptation of cut-off values). The evaluation of the performance is unbiased and not used to optimize model parameters.
Performance results consistency	Model performance obtained in the training, validation and testing sets is reported. Consistency checks of performance measures across the different sets are performed.
Comparison to conventional metrics	Performance of radiomics-based models is compared against conventional metrics such as tumor volume and clinical variables (e.g., staging) in order to evaluate the added value of radiomics (e.g., by assessing the significance of AUC increase calculated with the DeLong test).
Multivariable analysis with non-radiomics variables	Multivariable analysis integrates variables other than radiomics features (e.g., clinical information, demographic data, panomics, etc.).
CLINICAL IMPLICATIONS	
Biological correlate	Assessment of the relationship between macroscopic tumor phenotype(s) described with radiomics and the underlying microscopic tumor biology.
Potential clinical application	The study discusses the current and potential application(s) of proposed radiomics-based models in the clinical setting.
MATERIAL AVAILABILITY	
Open data	Imaging data, tumor ROI and clinical information are made available.
Open code	All software code related to computation of features, statistical analysis and machine learning, and allowing to exactly reproduce results, is open source. This code package is ideally shared in the form of easy-to-run organized scripts pointing to other relevant pieces of code, along with useful sets of instructions.
Open models	Complete models are available, including model parameters and cut-off values.