Responsible Radiomics Research for Faster Clinical Translation

Martin Vallières[1], Alex Zwanenburg[2,3], Bodgan Badic[1], Catherine Cheze Le Rest[1], Dimitris Visvikis[1], and Mathieu Hatt[1]

[1]*LaTIM, INSERM, UMR 1101, IBSAM, UBO, UBL, Brest, France;* [2]*National Center for Tumor Diseases, Dresden, Germany; and* [3]*German Cancer Research Center, Heidelberg, Germany*

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, because radiologic images are routinely acquired during cancer management. Imaging modalities such as ^{18}F-FDG PET, CT, and 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 consists mostly of visual or manual assessments. Today, the development of machine-learning techniques and the rise of computational power allow for the exploitation of a large number of quantitative features (*4*). This ability has led to a new incarnation of computer-aided diagnosis, "radiomics," which refers to the characterization of tumor phenotypes via the extraction of high-dimensional mineable data—for example, morphologic, intensity-based, fractal-based, and textural features—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 used in 2010 to describe how imaging features can reflect gene expression (*6*). Other early radiomics studies followed (*7,8*), including some highlighting early on that the reliability of existing features is affected by acquisition protocol, reconstruction, test–retest consistency, preprocessing, and segmentation (*9–13*). The overall framework of radiomics was then explicitly described in 2012 (*14*), and in the years that followed, this emerging field experienced 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): medical image acquisition, computation of radiomics features, and statistical analysis and machine learning. To apply the models to new patients for treatment personalization, a prospective model evaluation (preferably in a multicenter setup) is necessary.

Radiomics research has already shown great promise for supporting clinical decision making. However, the fact that radiomics-based strategies have not yet been translated to routine practice can be partly attributed to the low reproducibility of most current studies. The workflow for computing features is complex and involves many steps (Fig. 1B), often leading to incomplete reporting of methodologic information (e.g., texture matrix design choices and gray-level discretization methods). As a consequence, 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, the high false-positive rates (similar to those of analogous computer-aided diagnosis methods), and the reporting of overly optimistic results, all of which affect the generalizability of the conclusions reached in current studies.

Medical imaging journals are currently overwhelmed by a large volume of radiomics-related articles of variable quality and associated clinical value. The aim of this editorial is to present guidelines that we think can improve the reporting quality and therefore the reproducibility of radiomics studies, as well as the statistical quality of radiomics analyses. These guidelines can serve not only the authors of such studies but also the reviewers who assess their appropriateness for publication.

GUIDELINES FOR IMPROVING QUALITY OF RADIOMICS ANALYSES

The complexity of the radiomics workflow increases the need to standardize computation methods (*16–19*). Since September 2016, about 55 researchers from 19 institutions in 8 countries have participated in the Image Biomarker Standardization Initiative (IBSI), which aims at standardizing both the computation of features and the image-processing steps required before feature extraction (e.g., image interpolation and discretization). First, a simple digital phantom with few discrete image intensities was used to standardize the computation of 172 features from 11 categories. Then, 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 on 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, can serve as a calibration tool for future investigations. Ultimately, it may 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.
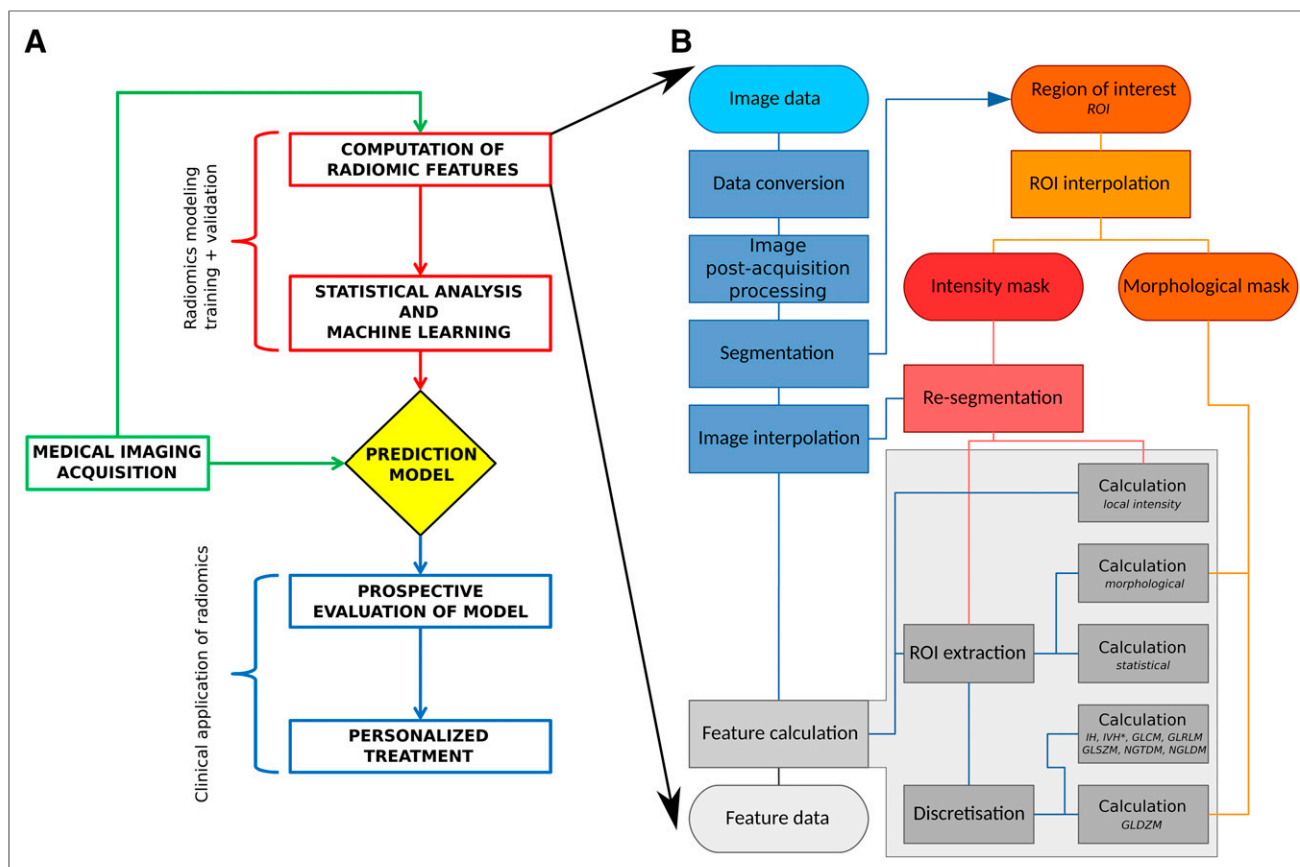
FIGURE 1. Radiomics workflow. (A) From medical imaging acquisition to treatment personalization. (B) Workflow of computation of radiomics features. Depending on the specific imaging modality and purpose, some steps may be omitted. In the figure, the feature calculation part is expanded to show different feature families with specific processing steps. IH = intensity histogram; IVH = intensity-volume histogram; GLCM = grey level cooccurrence matrix; GLRLM = grey level run length matrix; GLSZM = grey level size zone matrix; NGTDM = neighborhood grey tone difference matrix; NGLDM = neighboring grey level dependence matrix; GLDZM = grey level distance zone matrix. *Discretisation of IVH differs from IH and textural features. (Adapted from (20); ©2016-2017 IBSI. Creative Commons Attribution 4.0 International License.)

It would also be desirable that the code of existing software be updated to conform with 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 methodologic details, including the comprehensive description of image acquisition protocols, sequence of operations, image postacquisition processing, tumor segmentation, image interpolation, image resegmentation 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.

After 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 holds great potential, may be of interest if compared with the most reproducible and robust features or prognostic clinical information already used. Nonetheless, for the construction of prediction models via multivariable analysis, there are two basic requirements. First, all methodologic details and clinical information must be clearly reported or described to facilitate reproducibility and comparison with other studies and metaanalyses. Second, radiomics-based models must be tested in sufficiently large patient datasets distinct from teaching (training

and validation) sets to statistically demonstrate their efficacy over conventional models (e.g., existing biomarkers, tumor volume, and cancer stage). Ideally, for optimal reproducibility potential, all data and programming code related to the study should also be made available to the community. Table 2 provides guidelines based on the "radiomics quality score" (www.radiomics.world), which can help evaluate the quality of radiomics studies. More guidelines on reproducible prognostic modeling can be found in the TRIPOD statement (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) (22).

RESPONSIBLE RESEARCH IS THE KEY

Some guiding principles already exist to help radiomics scientists further implement the responsible research paradigm into their current practice. For one, the Responsible Research and Innovation website (www.rri-tools.eu) provides useful guidelines. Furthermore, a concise set of principles for better scientific data management and stewardship—the "FAIR guiding principles" (23)—has been defined, stating that all research objects should be findable, accessible, interoperable, and reusable. 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

TABLE 1
Reporting Guidelines on Computation of Radiomics Features

Category	Guideline.
General	
Image acquisition	Acquisition protocols and scanner parameters such as 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).
Volumetric analysis	Specification of whether imaging volumes were analyzed as separate images (2-dimensional) or as fully-connected volumes (3-dimensional).
Workflow structure	Sequence of processing steps leading to extraction of features.
Software	Software type and version of code used for computation of features.
Image preprocessing	
Conversion	How data were converted from input images (e.g., conversion of PET activity counts to SUV and calculation of ADC maps from raw diffusion-weighted MRI signal).
Processing	Image-processing steps after acquisition (e.g., noise filtering, intensity nonuniformity correction in MRI, and partial-volume effect corrections).
ROI segmentation*†	How ROIs were delineated in images (e.g., software or algorithms used, number of persons and their level of expertise [specialty, experience], method of reaching consensus, and mode [automatic or semiautomatic]).
Interpolation	
Voxel dimensions	Original and interpolated voxel dimensions.
Image interpolation method	Method used for interpolating voxel values (e.g., linear, cubic, or spline) and for aligning original and interpolated grids.
Intensity rounding	Rounding procedures for noninteger interpolated gray levels (if applicable) (e.g., rounding of Hounsfield units in CT images after interpolation).
ROI interpolation method	Methods used for interpolating ROI masks and for aligning original and interpolated grids.
ROI partial volume	Minimum partial-volume fraction required to include an interpolated mask voxel in the interpolated ROI (if applicable) (e.g., minimum partial-volume fraction of 0.5 when using linear interpolation).
ROI resegmentation	
Inclusion/exclusion criteria	Criteria for inclusion or exclusion of voxels from the ROI intensity mask (if applicable) (e.g., exclusion of voxels with Hounsfield unit values outside predefined range inside the ROI intensity mask on CT images).
Image discretization	
Discretization method	Method used for discretizing image intensities before feature extraction (e.g., fixed bin number, fixed bin width, and histogram equalization).
Discretization parameters	Parameters for image discretization (e.g., number of bins, bin width, and minimal value of discretization range).
Feature calculation	
Feature set	Description and formulas of all calculated features.
Feature parameters	Settings for calculation of features (e.g., voxel connectivity, with or without merging by slice, and with or without merging directional texture matrices).
Calibration	
Image-processing steps	Specification of which image-processing steps match benchmarks of the IBSI.
Feature calculation	Specification of which feature calculations match benchmarks of the IBSI.

*To reduce interobserver variability, automatic and semiautomatic methods are favored.
†In multimodal applications (e.g., PET/CT or PET/MRI), ROI definition may involve propagation of contours between modalities via coregistration. In that case, technical details of registration should also be provided.
ROI = region of interest.

TABLE 2
Quality Factors in Radiomics Studies

Factor	Description
Imaging	
Standardized imaging protocols	Imaging acquisition protocols are well described and ideally similar across patients. Alternatively, methodologic steps are taken toward standardizing them.
Imaging quality assurance	Methodologic steps are taken to incorporate only acquired images of sufficient quality.
Calibration	Computation of radiomics features and image-processing steps matches benchmarks of the IBSI.
Experimental setup	
Multiinstitutional/external datasets	Model construction or performance is evaluated using cohorts from different institutions, ideally from different parts of world.
Registration of prospective study	Prospective studies provide the highest level of evidence supporting clinical validity and usefulness of radiomics models.
Feature selection	
Feature robustness	Robustness of features is evaluated against segmentation variations and varying imaging settings (e.g., noise fluctuations and interscanner differences). Unreliable features are discarded.
Feature complementarity	Intercorrelation of features is evaluated; redundant features are discarded.
Model assessment	
False-discovery corrections	Corrections for multiple testing comparisons (e.g., Bonferroni or Benjamini–Hochberg) is applied in univariate analysis.
Estimation of model performance	Teaching dataset is separated into training and validation sets to estimate optimal model parameters (e.g., using bootstrapping, cross-validation, and random subsampling)
Independent testing	A testing set distinct from the teaching set is used to evaluate performance of complete models (i.e., without retraining and without adaptation of cutoffs). Evaluation of performance is unbiased and not used to optimize model parameters.
Performance results consistency	Model performance in 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) to evaluate the added value of radiomics (e.g., by assessing the significance of AUC increase with the DeLong test).
Multivariable analysis with nonradiomics variables	Multivariable analysis integrates variables other than radiomics features (e.g., clinical information, demographic data, and panomics).
Clinical implications	
Biologic correlate	Relationship between macroscopic tumor phenotypes described with radiomics and underlying microscopic tumor biology is assessed.
Potential clinical application	Current and potential applications of proposed radiomics-based models in clinical setting are discussed.
Material availability	
Open data	Imaging data, tumor region of interest, and clinical information are made available.
Open code	Software code for computation of features, statistical analysis, machine learning, and exact reproduction of results, is open-source. Code package is ideally shared as 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 cutoffs.

AUC = area under the receiver-operating-characteristic curve.

we can add two other tools of interest: the Cancer Imaging Archive (www.cancerimagingarchive.net), a service that anonymizes and hosts medical images for public download, and the Radiomics Ontology (www.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 radiomics data and methods, and would more concisely summarize the implementation details of a given radiomics workflow (e.g., Table 1).

To conclude, initial pioneer studies in radiomics have paved the way to an exciting field and to most promising methods for better personalizing 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. The best way to reach this goal is through responsible radiomics research, which can be summarized into three working principles that we should all try to follow as a research community: design and conduct high-quality radiomics research, write and present fully transparent radiomics research, and share data and methods.

DISCLOSURE

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 ^{18}F-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 ^{18}F-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 ^{18}F-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 (Bellingham).* 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. *arXiv1612.07003.* 2016.
21. Zwanenburg A. EP-1677: multicentre initiative for standardisation of image biomarkers [abstract]. *Radiother Oncol.* 2017;123(suppl):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.