Radiomics in PET/CT: Current Status and Future AI-Based Evolutions

Mathieu Hatt, Catherine Cheze Le Rest, Nils Antonorsi, Florent Tixier, Olena Tankyevych, Vincent Jaouen, Francois Lucia, Vincent Bourbonne, Ulrike Schick, Bogdan Badic, et al.

Radiomics in PET/CT: current status and future AI based evolutions

M. Hatt[1], C. Cheze Le Rest[1,2], N. Antonorsi[2], F. Tixier[1], O. Tankyevych[2], V. Jaouen[1,3], F Lucia[1], V Bourbonne[1], U. Schick[1], B Badic[1], D. Visvikis[1]

[1]LaTIM, INSERM, UMR 1101, University of Brest, CHRU Brest, France

[2]Nuclear Medicine Department, CHU Milétrie, Poitiers, France

[3]IMT-Atlantique, Plouzané, France

Keywords: radiomics; nuclear medicine, machine learning; deep learning

No financial supports

Corresponding author: Dimitris Visvikis

LaTIM, INSERM, UMR 1101

IBRBS, Faculty of Medicine

22 rue Camille Desmoulins

29238 Brest

Tel: +33 2 98 01 81 14

E-mail: dimitris@univ-brest.fr

No potential conflicts of interest relevant to this article exist.

Wordcount: ~7700

ABSTRACT

This short review aims at providing the readers with an update on the current status, as well as future perspectives in the quickly evolving field of radiomics applied to the field of PET/CT imaging. Numerous pitfalls have been identified in study design, data acquisition, segmentation, features calculation and modeling by the radiomics community, and these are often the same issues across all image modalities and clinical applications, however some of these are specific to PET/CT (and SPECT/CT) imaging and therefore the present paper focuses on those. In most cases, recommendations and potential methodological solutions do exist and should therefore be followed to improve the overall quality and reproducibility of published studies. In terms of future evolutions, the techniques from the larger field of artificial intelligence (AI), including those relying on deep neural networks (also known as deep learning) have already shown impressive potential to provide solutions, especially in terms of automation, but also to maybe fully replace the tools the radiomics community has been using until now in order to build the usual radiomics workflow. Some important challenges remain to be addressed before the full impact of AI may be realized but overall the field has made striking advances over the last few years and it is expected advances will continue at a rapid pace.

Introduction

Nuclear medicine imaging has been relying on positron emission tomography / computed tomography (PET/CT) combined scanners for 20 years [1]. Hardware and software innovations have allowed improved spatial resolution and signal-to-noise ratio of quantitative PET images, thanks to reconstruction algorithms able to integrate information regarding the specific point-spread-function and time-of-flight characteristics of the scanner [2,3]. Despite these hardware and algorithmic achievements, PET/CT images have been and still are mostly exploited visually in routine clinical practice, and using only simple standard metrics (such as SUV_{max} or SUV_{peak}, the maximum intensity voxel and the aggregate of several high intensity voxels, respectively) in clinical trials and publications [4]. The use of SUV maximum intensity has shown diagnostic and staging capabilities in several clinical settings, although thresholds and actual recommended values vary greatly across studies and meta-analyses. However, its limited discriminative power in baseline prognosis [5] or therapy response evaluation and prediction [6] has also been highlighted. This prompted various developments during the early 2000's, aiming at extracting more complete information from PET and SPECT images, with the hope that more complete characterization of lesions beyond simplistic measurements such as SUV_{max}, could provide more accurate and reliable information regarding diagnosis and prognosis for personalized patient management. Since the early 2010's, this field has been denoted "radiomics". Radiomics can be described as the extraction of numerous ("high-throughput") quantitative metrics (handcrafted and/or more recently obtained from pre-trained deep networks) from medical images, exploited to build decision-support models (e.g., diagnosis or prognosis). The standard workflow of radiomics consists of i) study design, ii) collection and curation of images and associated contextual data, iii) identification and delineation of objects of interest (e.g., organs, tumors), iv) extraction of features (intensity, shape, textural, "deep") and v) modeling (i.e., train and validate multiparametric models for a specific endpoint).

The goal of this short review is to provide an update on the current status and future perspectives of radiomics in PET/CT with a particular focus in the emerging role of artificial intelligence [7,8].

A short history of radiomics use in nuclear medicine imaging

Although the "radiomics" term was initially coined within the context of radiotherapy applications and computed tomography (CT) imaging relied upon for treatment planning in 2012 [9,10], similar studies had been carried out in PET as soon as the early 2000's, investigating the potential value of handcrafted features extracted from PET images [11–13]. One of the first studies developed a handcrafted metric to quantify heterogeneity in the FDG uptake and applied it to PET images of sarcoma patients, showing improved prognostic value of the heterogeneity metric compared to SUV_{max} [12]. A later study investigated the use of shape, intensity and textural features (most of what constitutes today the cornerstone of radiomic features) in PET images to predict outcome as a proof of concept in a very small number of patients [11]. In 2010, another study investigated the impact of changes in acquisition protocols and reconstruction settings on the derived intensity and textural features [13]. In 2011, a study evaluated the value of some textural features calculated from FDG PET images to predict response to therapy in esophageal cancer patients [14]. And in 2012, the reproducibility of these heterogeneity metrics was evaluated in test-retest PET images [15]. When the term "radiomics" began to be adopted by a growing community after 2012 and the paper by Lambin et al, publications applying that concept in nuclear medicine images started to use that keyword as well. Since then, the number of studies applying the radiomics concept to PET or SPECT images and using that denomination has been steadily increasing. By the middle of the year 2020, >1100 publications (excluding letters, editorials, abstracts and meetings) using the term "radiomics" can be found in the web of science databases. Only 16% of them use also the terms PET or PET/CT,

and only a few SPECT/CT (e.g., [16]). Today, publications regarding the use of radiomics in PET/CT are numerous and examples can be found for most of the usual cancer types (*i.e.*, lung, breast, brain, head and neck, rectum, etc.). In majority most of the works in PET/CT have been largely concentrated on ^{18}FDG imaging, albeit a few exceptions [17]. However, these studies share most of the pitfalls and limitations observed for radiomics studies in general: small, retrospective and monocentric cohorts, in certain cases flaws in statistical analysis, and limited reproducibility/external validation of the findings. The quality and reliability of most studies has been moderate at best [18], which prevented a quick translation to the clinical practice, despite some promising results. Over the last few years however, a positive trend can be observed, with studies showing increasing levels of quality, reliability and reproducibility, on larger cohorts, with more robust statistical analysis and modeling strategies relying on machine learning techniques and rules, including the use of external validation, hence a more convincing level of evidence. This evolution was encouraged and supported by a number of publications and recommendation initiatives aiming at the improvement of practices, standardization and reproducibility [19–26]. In parallel to these improvements, the radiomics community has also started relying on deep learning (DL) [8, 27–30], which constitutes one of the main perspectives in the field to solve some of the issues and limitations of the current radiomics workflow as described in the introduction, such as for instance an improved automation of the whole process. However, the use of DL methods in radiomics obviously also requires solving new issues and faces numerous challenges, like the high dependency on large datasets and the limited interpretability of the resulting networks.

Main pitfalls and limitations of current PET/CT radiomics studies

1. *Poor study design and data collection*

A specific example is to start a radiomic study relying on PET/CT images without having datasets of sufficient size and of different sources in order to satisfy the training/validation/testing requirements [31]. Most of the published studies to date have been monocentric (unlikely to be generalizable to other sites), retrospective (selection bias), on cohorts with size between 50 and 100 patients (risks of overfitting for training models, difficult to rigorously cross-validate). Irrespectively of the quality of the rest of the analysis, such characteristics of a study constitute a strong limitation that is almost impossible to overcome in order to provide convincing results with a high level of confidence. The resulting trained/cross-validated models are unlikely to be well validated in future external data.

Images are collected retrospectively and/or acquired prospectively for the purpose of carrying out radiomic analyses. If PET images are retrospectively collected, the associated raw data are usually not available and therefore the reconstructed images have to be exploited as they are, although they can be processed to be improved to a certain extent, for denoising or partial volume effects correction purposes. On the contrary, if images are acquired prospectively, it is recommended that raw data are stored for research purposes, such as exploring alternative image reconstruction settings that may be beneficial for radiomics analyses [32].

Another issue is related to the collection of additional contextual data from clinical records, for which curation quality checks are crucial. These data are usually extracted by investigators in the medical records and entered in specific research databases in an entirely manual way, which is obviously prone to errors such as duplicates, missing data or falsely attributed labels. These can be very difficult to detect and identify later in the process, which is why designing robust data infrastructures is one of the needs to address in the future [23]. In addition, it will be appropriate to develop methodologies that can handle such heterogeneous in terms of completeness datasets.

2. *Insufficient reporting*

Most of published studies do not provide sufficient details to allow for reproduction of their findings in similar datasets. Current recommendations are quite exhaustive regarding the exact details that should be reported fully and in supplemental materials of publications, and the field has clearly made some progress in that regard over the last few years [19,33]. However, actual data sharing is also very rare in the field, which in practice prevents re-analysis of the exact same dataset for further replication, as well as external validation of the developed signatures and models. Finally, as in other fields, there exists a bias to publish and report more easily positive results than negative ones [34]. In the future, authors need to avoid self-censorship and willingly report negative results or radiomic failures and editors and reviewers need to consider favorably such studies for publication if the methodology is sound and the results support the conclusions.

3. Lack of standards

The calculation of most radiomic features involves several steps. A number of different pre-processing choices have to be made, especially for textural features. Their implementation is thus prone to errors but also to very different outputs despite relying on the exact same definition and formulae.

Fortunately, this has much improved over the last few years thanks to the Imaging Biomarkers Standardization Initiative (IBSI) [20], carried out by more than 20 research groups from 8 different countries. The IBSI has established standardized definitions and nomenclature of 172 handcrafted features usually considered, as well as a standard nomenclature for the full radiomics pipeline and each step of pre- and post-processing leading to features extraction. The IBSI also established recommendations specifically for the various important steps and details such as pixel/voxel interpolation, intensity discretization and texture matrices design. Finally, a benchmark of radiomic features values on both synthetic digital phantom and real clinical images for each radiomic feature calculated in different possible configurations is now available. Most of the IBSI recommendations, guidelines and results (reference document updated 03/2020 available online[1]) are directly applicable to PET radiomics. For instance, it is highly recommended to check the IBSI compliance of homemade or commercial/open-source libraries and software before considering using them in an analysis, as this will greatly improve the reproducibility of the findings.

4. Segmentation

Another area in the current radiomics process, where the lack of full automation and a clear standard is especially hindering the widespread reproducibility and acceptability of radiomics in PET, is the need for detection and delineation of objects of interest such as tumors before features can be extracted. Indeed, even if an otherwise perfectly standardized workflow of radiomic feature extraction is established, the use of different volumes of interest resulting from the use of different segmentation methods will prevent the results to be perfectly reproduced. The impact of the variations in segmentation has been investigated for PET radiomics and can be significant in terms of the resulting value of extracted features. The task of automated delineation of volumes of interest has been the focus of numerous methodological developments since the end of the 90's, first focusing on very basic thresholding approaches, until the recently applied convolutional neural networks (CNN) based on U-Net architecture [35]. The task is even more complex when multiple lesions need to be analyzed, and until recently, most semi-automated or fully automated techniques assumed the lesion to segment had been isolated in a volume of interest first, which obviously relies

[1] https://ibsi.readthedocs.io/en/latest/

most of the time on user intervention, hence a lack of full automation to process large cohorts for a radiomic analysis. This makes the segmentation step the most time-consuming bottleneck of the radiomic workflow in most PET radiomic studies. Current recommendations are to rely on (semi)automated methods as much as possible, avoid basic fixed thresholding [35-36], and potentially consider a consensus of several techniques [37] or methods identifying the most appropriate algorithm for a given image [38]. The current state of the art methods for achieving fully automated PET image segmentation are based on deep neural networks [36, 39-40], which have been very successful in medical image segmentation tasks [41-42]. As the learning process relies on pixels/voxels or patches, the amount of data necessary for learning is relatively small for an efficient training. DL-based methods integrating objects of interest detection and segmentation [43] may facilitate the full automation of this step of the radiomics pipeline, allowing for radiomic analyses of hundreds or thousands of patients datasets in a more convenient and less time-consuming fashion.

5. Multicenter harmonization

One of the most important limitations of PET radiomics studies has been the use of small, monocentric cohorts, thus leading to a low level of evidence, with developed models or signatures almost never being evaluated in external datasets [44]. In order to train more generalizable models and increase the statistical relevance of findings, the use of much larger (i.e., several hundreds) cohorts of patients is warranted [22,45]. This however raises a number of challenges, including data sharing legal, ethical, administrative and technical issues. Sharing and analyzing large multicentric cohorts of patients is clearly not today's reality of radiomic studies. However, even if this is solved, thanks for example to distributed learning [46], which provides a way to train the models in each institution without the need to actually move the data out of each center, an additional issue is hampering such analyses. Indeed, radiomic features are notoriously sensitive to variations in scanner model and manufacturer, acquisition protocols and reconstruction settings [13,47]. It is thus ill-advised to pool features from patients with images having different characteristics, since the biases and shifts in features values and distributions could either mask true correlation with outcome or create false positive relationships in the statistical analysis [48]. A number of possible methodological approaches have been devised recently to address this problem. First, it can be mitigated to some extent by improving standardization of acquisition and reconstruction protocols settings in case of prospective multi-center data collection, as well developed guidelines exist for PET/CT imaging [49,50]. This however is only feasible for prospectively acquired data, whereas most radiomic studies are still performed by collecting existing data retrospectively, where it is not possible to modify the acquisition and reconstruction settings anymore. In addition, the standardization guidelines focus on SUV standardization, which is insufficient to address all image characteristic changes that have significant impact on some radiomic features [51,52]. Other pre-processing steps, such as interpolation to a common voxel size and/or filtering methods, could help in reducing the differences across images. Such an approach however could also introduce artifacts or reduce the quality of the quantitative information contained in the images.

Selection of features sufficiently robust or even completely insensitive to the variability of scanner, acquisition and/or reconstruction settings, can obviously help building robust models. Most investigations on that issue in PET imaging have shown that reliability of features (either robustness or repeatability and test-retest reliability) [13,15,53–55] is very much variable among features, as well as within each category of features. It can therefore be challenging to identify a threshold characterizing features robust enough to be kept for further analysis. Another drawback is that numerous potentially informative features are removed before being evaluated, which leads to an important loss of extracted information. The most recently proposed and evaluated method is to

perform the harmonization in the feature space, using statistical normalization and analysis techniques. A recent review provides an overview of the methods that have been proposed to achieve this, and the subset that found their way to radiomics applications [56]. One of the promising methods is ComBat, which was initially developed to correct for batch effects in genomics studies [57] and was shown to be efficient with small sample sizes and to outperform similar methods [58]. ComBat was applied in PET radiomics, showing improvement in the resulting external validation of the developed models [59-60]. There are several advantages in the use of ComBat. It is much easier than processing images and it is very fast to apply. It allows relying on the entire feature space that was extracted, thus no loss of information. Recently, an improved version of ComBat relying on the addition of bootstraping in the estimation step was shown to slightly but consistently improve the external validation performance of predictive models [61]. One important limitation of ComBat is that at least a sample of each of the different center datasets need to be available and labeled. Also, it cannot be easily applied on an individual patient basis or directly to new, previously unseen data. Other approaches including re-scaling and normalization have been recently evaluated with interesting results to improve multi-center modeling [62-63]. Although some of these recent developments were applied to CT or MRI imaging, their results are directly transferable to PET imaging.

Emerging DL based techniques could provide a robust solution allowing to work in the image space rather than in the radiomics space; hence harmonising images rather than radiomics features. Potential advantages include the possibility to automate the whole radiomics process from images to modeling as well as facilitating model usage for single patients rather than patient cohorts which are necessary with the current harmonization techniques in the radiomics space. Methods based on generative adversarial networks (GAN) could help in generating images with appropriate properties while preserving the clinically-relevant information they contain. A recent study showed in CT that it was possible to generate images through DL from a given kernel to another, which eliminated most of the bias in the resulting extracted radiomic features [64]. Such an approach could easily be translated to PET.

6. Volume and other metrics confounding issues

As radiomic features include usual PET measurements (*i.e.,* functional metabolic volume corresponds to the geometrical volume feature, SUV_{mean} or SUV_{max} correspond to mean and max intensities respectively, etc.), it is especially crucial to check redundancy and complementary value of any additional feature that is investigated. It is indeed useless to calculate a complex new feature designed to quantify uptake heterogeneity or tumor shape if this feature ends up being very highly correlated with volume or SUV_{max} [53], as it will therefore provide little to no additional value.

This has been an especially pervasive issue for metabolic tumor volume in PET/CT radiomics studies, (mostly those published before 2015). Because all features are derived from an already determined volume of interest and most texture definitions are dependent on the number of voxels involved in the calculation, numerous textural features were found to be highly correlated with volume. Similarly, most of shape descriptors are naturally dependent on volume (e.g., maximum diameter and asphericity are proportional to tumor volume, larger tumors exhibiting less spherical shapes and obviously larger diameters). Overall, it was shown that most choices in the calculation of features, including the discretization method or matrix design had a strong impact on feature values, statistical distributions and correlation with other metrics [45,54,55,65,66]. For instance, it was suggested that alternative intensity discretization schemes such as fixed bin width instead of the at the time mostly used fixed bin number, led to textural features with lower redundancy with the corresponding volume [67]. However, it was also shown that features then became more correlated with maximum

intensity [45,67]. It was also claimed that no such added value could be obtained for volumes below 45cm^3 [68], which was indeed the case of some previous PET radiomic studies, where the identified textural features were actually surrogates of tumor volumes [53]. However this analysis was based on a single textural feature, calculated with very specific formulation choices and it was later shown alternative implementations could significantly improve the complementary value provided by the same feature and others demonstrating complementary value between tumor functional volume and texture analysis in a wide range of cancer models for tumor volumes >10cm^3 [55]. Clearly, the challenge is being able to estimate this threshold below which a metric can be expected to provide complementary value to the corresponding volume or is simply too redundant to provide additional information. This is why it is recommended to always include multivariate analysis (i.e., take into account redundancy and intercorrelation of extracted features), as well as to systematically compare the performance of any proposed radiomic-based model to "clinical only" and "volume and/or SUV only" models. Note that similar warnings were published recently regarding the CT-derived radiomics signature previously identified by Aerts et al. [23,69].

7. New features and modeling

Given the very large spectrum of calculation choices, parameters and methods to derive a value for a single textural parameter, it can be possible to have dozens of different values. This can be considered as optimizing textural analysis [32,66]. Indeed, it may be necessary to rely on different parameters to optimize the resulting discriminative power of each radiomic feature regarding a specific endpoint. It may therefore be beneficial to extract features with a large range of possible pre-processing choices, intensity discretization methods and settings, or even textural matrices designs and merging strategies. However on the other hand, such an approach has the drawback of artificially inflating the size of a dataset, hence requiring a very robust and reliable feature selection and machine learning pipeline in order to handle the very large resulting set of variables. Beyond the usual set of imagie derived features used in most radiomic studies and that have been standardized by the IBSI, some studies have also proposed alternative and new handcrafted features. These may have higher discriminant power or improved properties. CoLIAGe (Co-occurrence of Local Anisotropic Gradient Orientations) [70], metabolic gradient [71] or 3-D Riesz-covariance textures [72] are recent examples that were shown to have higher differentiation power compared to the usual textural features. Specifically for PET, a novel metric was designed as an alternative to textures with the goal of quantifying PET heterogeneity by yielding increased values for tumors with peripheral sub-regions of high SUV [73]. Similarly, a new GLCM methodology was recently proposed to reduce redundancy amongst calculated features and was shown to allow for higher accuracy in classifiying tumor types in CT [74]. Obviously, DL techniques have also provided numerous "new" features, such as those extracted by medical images using deep networks pre-trained on other datasets to extract "rough" to "fine" features at different scales through different layers [75]. These known as "deep features" can be exploited directly as well as combined with other handcrafted radiomic features to build better models [76–80].

The last step of the whole radiomic workflow consists in building models (diagnostic, predictive, prognostic) for a given clinical endpoint or task, by relying on the set of features that has been extracted from the images, combined with other available data such as clinical information or histopathological and other –omics variables. This type of studies represent the weakest part of the literature in PET radiomics studies until recently, with mostly inappropriate statistical analysis prone to overfitting, false positives and overoptimistic claims, because of the lack of appropriate multivariate analysis, lack of corrections for false discovery, small sample sizes (most studies with less than 150 patients [60]) and lack of external validation [33,45,81]. However, a proper methodology

relying on splitting the data into training, validation and testing datasets, the use of cross validation and proper metrics for optimizing hyperparameters and evaluating the models, and handling the imbalance in the data, may not be sufficient to alleviate all risks. Indeed, even the choice of the feature selection and classifier can lead to significantly different resulting performance, as the most recent comparison papers have highlighted [82–84]. This suggests that implementing several different techniques and relying on their consensus could improve the prediction performance [85]. Finally, DL techniques, again may provide alternative solutions to the usual modeling step in the radiomics workflow [86-89]. For instance, deep neural networks can achieve feature selection and classification, but it can also be envisioned that one or several neural networks can replace in the future parts or all of the radiomics workflow [7].

Perspectives of PET radiomics

As an overall conclusion one can easily state that the field of PET radiomics has gained a lot in terms of maturity over the last decade. Such improvements concern the use of: i) larger cohorts of patients, ii) proper machine learning and even deep learning methods to facilitate, automate the workflow and make modeling more robust and reliable, iii) more rigorous segmentation of regions of interest, iv) standardized features definition, nomenclature and reporting for improved reproducibility, v) validation of models in external datasets. However, the field has not yet been able to deliver the level of proof that is necessary for the clinical community to fully adopt the approach [18,90]. The community of radiomics, especially in the field of PET/CT imaging, needs to further strengthen its efforts to provide such convincing evidence as well as established benchmarks and standards for user-friendly software and guidelines. This includes the expansion and finalization of the standardization (e.g., the version 2 of the IBSI focusing on filter-based features currently ongoing), the development of tools and methods to facilitate the collection, storage and sharing of multicenter data, increasing the level of automation of the whole workflow and identify the appropriate steps for integrating DL methods in the radiomics workflow. Regarding the required level of evidence, the time seems right to carry out larger, multicentric and prospective studies. Finally, the requirement for explainable / interpretable models is crucial for radiomics, as multiparametric radiomic models derived by DL approaches based on complex features may be difficult for clinicians to interpret and therefore to trust [91-92]. Although this is already challenging for standard current state of the art radiomics, it will become even more of an issue when relying on deep neural networks, as these contain millions of parameters to optimize and can therefore be seen purely as “black boxes” impervious to human understanding. Methods for providing visual feedback through activation / saliency maps and network visualization techniques [93–95] will be of paramount importance to allow DL-based radiomic models being interpretable for end users and hence hopefully be accepted by clinicians in routine practice.

Conclusion

Radiomics have been extensively investigated in PET/CT imaging. A few seminal studies in the field predate the actual coinage of the “radiomics” denomination in the literature. It is a very active and promising field of research, however the quality and reliability of published studies until recently can be questioned. There is however and fortunately a clear trend towards improvement, as the radiomics community has become very much aware of these issues. Regarding the perspectives, there are two different aspects that are likely to evolve quite rapidly in the near future. First, larger, multicentric (possibly prospective) studies may become the norm (instead of the exception today) of radiomic analyses, as data availability and sharing improves and methods for efficient harmonization become more readily available. Second, relying on deep learning techniques to automate and improve the process (e.g., detection/segmentation) and alleviate current limitations of the current

radiomics workflow (e.g., more straightforward feature extraction and selection) should help in developing efficient decision support systems. This however will require specific developments to help training networks with relatively small amounts of data and to “open the black box” in order to provide interpretable models that end users can trust and will be willing to rely upon in clinical practice.

References

1. Beyer T, Townsend DW, Brun T, Kinahan PE, Charron M, Roddy R, et al. A combined PET/CT scanner for clinical oncology. J Nucl Med. 2000;41:1369–79.

2. Berg E, Cherry SR. Innovations in Instrumentation for Positron Emission Tomography. Semin Nucl Med. 2018;48:311–31.

3. Jones T, Townsend DW. History and future technical innovation in positron emission tomography. J Med Imaging. 2017;4:011013.

4. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1:122S-50S.

5. Bonomo P, Merlotti A, Olmetto E, Bianchi A, Desideri I, Bacigalupo A, et al. What is the prognostic impact of FDG PET in locally advanced head and neck squamous cell carcinoma treated with concomitant chemo-radiotherapy? A systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2018;45:2122–38.

6. Kwee RM. Prediction of tumor response to neoadjuvant therapy in patients with esophageal cancer with use of 18F FDG PET: a systematic review. Radiology. 2010;254:707–17.

7. Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. Eur J Nucl Med Mol Imaging. 2019 Dec;46(13):2630-2637.

8. M. Hatt, C. Parmar, J. Qi and I. El Naqa, "Machine (Deep) Learning Methods for Image Processing and Radiomics," in IEEE Transactions on Radiation and Plasma Medical Sciences, 2019, vol. 3, no. 2, pp. 104-108.

9. 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;65:517–21.

10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.

11. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit. 2009;42:1162–71.

12. O'Sullivan F, Roy S, Eary J. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. Biostatistics. 2003;4:433–48.

13. 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;49:1012–6.

14. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, 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;52:369–78.

15. Tixier F, Hatt M, Le Rest CC, 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.

16. Rahmim A, Huang P, Shenkov N, Fotouhi S, Davoodi-Bojd E, Lu L, et al. Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images. NeuroImage Clin. 2017;16:539–44.

17. M. Majdoub et al., "Prognostic Value of Head and Neck Tumor Proliferative Sphericity From 3'-Deoxy-3'-[18F] Fluorothymidine Positron Emission Tomography," in IEEE Transactions on Radiation and Plasma Medical Sciences, 2018, vol. 2, no. 1, pp. 33-40

18. Pinto Dos Santos D, Dietzel M, Baessler B. A decade of radiomics research: are images really data or just patterns in the noise? Eur Radiol. 2020;

19. Vallières M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible Radiomics Research for Faster Clinical Translation. J Nucl Med Off Publ Soc Nucl Med. 2018;59:189–93.

20. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;191145.

21. Aerts HJWL. Data Science in Radiology: A Path Forward. Clin Cancer Res. 2017;

22. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017;14:169–86.

23. Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: The need for safeguards. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2019;130:2–9.

24. Morin O, Vallières M, Jochems A, Woodruff HC, Valdes G, Braunstein SE, et al. A Deep Look Into the Future of Quantitative Imaging in Oncology: A Statement of Working Principles and Proposal for Change. Int J Radiat Oncol Biol Phys. 2018;102:1074–82.

25. 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.

26. Sanduleanu S, Woodruff HC, de Jong EEC, van Timmeren JE, Jochems A, Dubois L, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. Radiother Oncol. 2018;127:349–60.

27. Hatt M, Le Rest CC, Tixier F, Badic B, Schick U, Visvikis D. Radiomics: Data Are Also Images. J Nucl Med Off Publ Soc Nucl Med. 2019;60:38S-44S.

28. Diamant A, Chatterjee A, Vallières M, Shenouda G, Seuntjens J. Deep learning in head & neck cancer outcome prediction. Sci Rep. 2019;9:2764.

29. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS Med. 2018;15:e1002711.

30. A. Amyar, S. Ruan, I. Gardin, C. Chatelain, P. Decazes and R. Modzelewski, "3-D RPET-NET: Development of a 3-D PET Imaging Convolutional Neural Network for Radiomics Analysis and Outcome Prediction," in IEEE Transactions on Radiation and Plasma Medical Sciences, 2019, vol. 3, no. 2, pp. 225-231

31. Chicco D. Ten quick tips for machine learning in computational biology. BioData Min. 2017;10:35.

32. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol. 2015;60:5471–96.

33. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging. 2019;46:2638–55.

34. Buvat I, Orlhac F. The Dark Side of Radiomics: On the Paramount Importance of Publishing Negative Results. J Nucl Med Off Publ Soc Nucl Med. 2019;60:1543–4.

35. Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, De Bernardi E, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. Med Phys. 2017;44:e1–42.

36. Hatt M, Laurent B, Ouahabi A, Fayad H, Tan S, Li L, et al. The first MICCAI challenge on PET tumor segmentation. Med Image Anal. 2018;44:177–95.

37. McGurk RJ, Bowsher J, Lee JA, Das SK. Combining multiple FDG-PET radiotherapy target segmentation methods to reduce the effect of variable performance of individual segmentation methods. Med Phys. 2013;40:042501.

38. Berthon B, Marshall C, Evans M, Spezi E. ATLAAS: an automatic decision tree-based learning algorithm for advanced image segmentation in positron emission tomography. Phys Med Biol. 2016;61:4855–69.

39. Zhao X, Li L, Lu W, Tan S. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. Phys Med Biol. 2018;

40. Zhong Z, Kim Y, Plichta K, Bryan GA, Zhou L, Buatti J, et al. Simultaneous Co-segmentation of Tumors in PET-CT Images Using Deep Fully Convolutional Networks. Med Phys. 2018;

41. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.

42. Z. Guo, X. Li, H. Huang, N. Guo and Q. Li, "Deep Learning-Based Image Segmentation on Multimodal Medical Imaging," IEEE Transactions on Radiation and Plasma Medical Sciences, 2019, vol. 3, no. 2, pp. 162-169

43. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PloS One. 2018;13:e0195798.

44. Zwanenburg A, Löck S. Why validation of prognostic models matters? Radiother Oncol. 2018;127:370–3.

45. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging. 2017;44:151–65.

46. Torres-Velázquez M, Chen WJ, Li X, McMillan A, Application and Construction of Deep Learning Networks in Medical Imaging, IEEE Transactions on Radiation and Plasma Medical Sciences, 2020; in press

47. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. J Nucl Med. 2015;56:1667–73.

48. Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, et al. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017;8:43169–79.

49. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54.

50. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA, et al. Feasibility of state of the art PET/CT systems performance harmonisation. Eur J Nucl Med Mol Imaging. 2018;45:1344–61.

51. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur J Nucl Med Mol Imaging. 2017;44:17–31.

52. Pfaehler E, van Sluis J, Merema BBJ, van Ooijen P, Berendsen RCM, van Velden FHP, et al. Experimental Multicenter and Multivendor Evaluation of the Performance of PET Radiomic Features Using 3-Dimensionally Printed Phantom Inserts. J Nucl Med Off Publ Soc Nucl Med. 2020;61:469–76.

53. 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;40:1662–71.

54. Desseroit M-C, Tixier F, Weber WA, Siegel BA, Cheze Le Rest C, Visvikis D, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort. J Nucl Med. 2017;58:406–11.

55. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi-Cancer Site Patient Cohort. J Nucl Med. 2015;56:38–44.

56. Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. Phys Med Biol. 2020; in press

57. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostat Oxf Engl. 2007;8:118–27.

58. Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, et al. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. PloS One. 2011;6:e17238.

59. Lucia F, Visvikis D, Vallières M, Desseroit M-C, Miranda O, Robin P, et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2019;

60. Orlhac F, Boughdad S, Philippe C, Stalla-Bourdillon H, Nioche C, Champion L, et al. A post-reconstruction harmonization method for multicenter radiomic studies in PET. J Nucl Med. 2018;

61. Da-Ano R, Masson I, Lucia F, Doré M, Robin P, Alfieri J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. Sci Rep. 2020;10:10248.

62. Chatterjee A, Vallières M, Dohan A, Levesque IR, Ueno Y, Saif S, et al. Creating robust predictive radiomic models for data from independent institutions using normalization. IEEE Trans Radiat Plasma Med Sci. 2019;1–1.

63. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. Sci Rep. 2020;10:12340.

64. Choe J, Lee SM, Do K-H, Lee G, Lee J-G, Lee SM, et al. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. Radiology. 2019;292:365–73.

65. Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, 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.

66. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep. 2017;7:10117.

67. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. PloS One. 2015;10:e0145063.

68. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med. 2014;55:37–42.

69. Vallieres M, Visvikis D, Hatt M. Dependency of a validated radiomics signature on tumor volume and potential corrections. J Nucl Med. 2018;59:640–640.

70. Prasanna P, Tiwari P, Madabhushi A. Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe): A new radiomics descriptor. Sci Rep. 2016;6:37241.

71. Wolsztynski E, O'Sullivan F, Keyes E, O'Sullivan J, Eary JF. Positron emission tomography-based assessment of metabolic gradient and other prognostic features in sarcoma. J Med Imaging. 2018;5:024502.

72. Cirujeda P, Dicente Cid Y, Muller H, Rubin D, Aguilera TA, Loo BW, et al. A 3-D Riesz-Covariance Texture Model for Prediction of Nodule Recurrence in Lung CT. IEEE Trans Med Imaging. 2016;35:2620–30.

73. Wang P, Xu W, Sun J, Yang C, Wang G, Sa Y, et al. A new assessment model for tumor heterogeneity analysis with [18]F-FDG PET images. EXCLI J. 2016;15:75–84.

74. Li X, Guindani M, Ng CS, Hobbs BP. Spatial Bayesian modeling of GLCM with application to malignant lesion characterization. J Appl Stat. 2019;46:230–46.

75. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis. 2015;115:211–52.

76. Paul R, Hawkins SH, Schabath MB, Gillies RJ, Hall LO, Goldgof DB. Predicting malignant nodules by fusing deep features with classical radiomics features. J Med Imaging. 2018;5:011021.

77. Ning Z, Luo J, Li Y, Han S, Feng Q, Xu Y, et al. Pattern Classification for Gastrointestinal Stromal Tumors by Integration of Radiomics and Deep Convolutional Features. IEEE J Biomed Health Inform. 2018;

78. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys. 2017;44:5162–71.

79. Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. Sci Rep. 2017;7:10353.

80. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep. 2017;7:5467.

81. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. PloS One. 2015;10:e0124165.

82. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5:13087.

83. Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. Sci Rep. 2017;7:13206.

84. Deist TM, Dankers FJWM, Valdes G, Wijsman R, Hsu I-C, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. Med Phys. 2018;45:3449–59.

85. Paul R, Hall L, Goldgof D, Schabath M, Gillies R. Predicting Nodule Malignancy using a CNN Ensemble Approach. Proc Int Jt Conf Neural Netw Int Jt Conf Neural Netw. 2018;2018.

86. Ypsilantis P-P, Siddique M, Sohn H-M, Davies A, Cook G, Goh V, et al. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. PloS One. 2015;10:e0137036.

87. Amyar A, Ruan S, Gardin I, Chatelain C, Decazes P, Modzelewski R. 3D RPET-NET: Development of a 3D PET Imaging Convolutional Neural Network for Radiomics Analysis and Outcome Prediction. IEEE Trans Radiat Plasma Med Sci. 2019;1–1.

88. Choi YS, Bae S, Chang JH, Kang S-G, Kim SH, Kim J, et al. Fully Automated Hybrid Approach to Predict the IDH Mutation Status of Gliomas via Deep Learning and Radiomics. Neuro-Oncol. 2020;

89. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun. 2020;11:1236.

90. Cheze Le Rest C, Hustinx R. Are radiomics ready for clinical prime-time in PET/CT imaging? Q J Nucl Med Mol Imaging Off Publ Ital Assoc Nucl Med AIMN Int Assoc Radiopharmacol IAR Sect Soc Of. 2019;63:347–54.

91. Hustinx R. Physician centred imaging interpretation is dying out - why should I be a nuclear medicine physician? Eur J Nucl Med Mol Imaging. 2019;46:2708–14.

92. Ibrahim A, Vallières M, Woodruff H, Primakov S, Beheshti M, Keek S, et al. Radiomics Analysis for Clinical Decision Support in Nuclear Medicine. Semin Nucl Med. 2019;49:438–49.

93. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding Neural Networks Through Deep Visualization. ArXiv150606579 Cs [Internet]. 2015 [cited 2018 Jan 26]; Available from: http://arxiv.org/abs/1506.06579

94. Brocki L, Chung NC. Concept Saliency Maps to Visualize Relevant Features in Deep Generative Models. ArXiv191013140 Cs Stat [Internet]. 2019 [cited 2020 Jun 13]; Available from: http://arxiv.org/abs/1910.13140

95. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE Int Conf Comput Vis ICCV. 2017. p. 618–26.