Radiomics in PET/CT: More Than Meets the Eye?

Mathieu Hatt, Florent Tixier, Dimitris Visvikis, Catherine Cheze Le Rest

Radiomics in PET/CT: more than meets the eye?

Mathieu Hatt[1], Florent Tixier[2], Dimitris Visvikis[1] and Catherine Cheze Le Rest[2]

1. LaTIM, INSERM, UMR 1101, University of Brest, IBSAM, Brest, France.
2. Academic department of Nuclear Medicine, CHU Poitiers, Poitiers, France.

Corresponding author: Mathieu Hatt

LaTIM, INSERM, UMR 1101

CHRU Morvan

2 avenue Foch

29609 Brest Cedex

Email: hatt@univ-brest.fr

Wordcount: ~1630

Radiomics is defined as the high throughput extraction of quantitative metrics from medical images (1). One of its main assumptions is that medical images are considered not merely as pictures for visual assessment, but rather as minable quantitative data (2), that may not necessarily be captured by the human eye ("more than meets the eye") (3).

In the present issue of the *J Nucl Med*, Orlhac and colleagues present a study comparing visual assessment of uptake heterogeneity in positron emission tomography (PET) images by experts and a subset of radiomics metrics, namely textural features (TF). They exploited both clinical and simple simulated PET images, going further than previous studies carried out using clinical data only (4–6). Such studies are useful because they provide additional understanding relative to the visual meaning of quantitative metrics that cannot be easily explained to non-specialists. These studies have focused on the PET component and the FDG uptake heterogeneity. Similar analyses have been carried out in computed tomography (CT) (7) or magnetic resonance imaging (MRI) (8).

One important finding is that TF calculated after a "relative" quantization process (*i.e.* resampling the original image intensities into a variable number of bins of fixed width, for example 0.5 standardized uptake values (SUV) (9)) correlate better with visual assessment than when calculated after the usual quantization process (*i.e.* uniformly resampling the original intensities into a set number of bins, for example 64 or 128). These different observations can be overall related to different factors, such as the very different correlative relationships between texture parameters and either the maximum intensity (SUV_{max}) and/or the number of voxels involved (tumour volume), that have been also previously reported (9–12). Other quantization processes (histogram equalization, Max-Loyd clustering, etc.) can lead to yet different distributions and associated clinical value (13).

The consensus amongst experts was also substantially higher than in earlier studies, mostly because only two categories (heterogeneous *vs.* homogeneous) were considered, compared to three (4), four (6) or even five (7) in previous studies. In (4), the visual assessment into three categories had limited prognostic value compared to TF (4). Since there was no clinical endpoint (survival, outcome) in the study by Orhlac, *et al*. we cannot conclude about the clinical value of the features that correlated well with visual assessment, although it is safe to assume that these features will be useful in clinical applications where there is a correlation between patients outcome and the level of uptake heterogeneity visually assessed (or SUV_{max}, given the observed correlations).

The primary goal of radiomics is to build clinical models using machine learning techniques (14) to predict patients' outcome, thereby allowing better patient management. These multiparametric models, which are likely to be unintelligible even to experts since they combine a large number of high-order multimodality image features (12,15), should outperform visual analysis in terms of both accuracy and reproducibility. To associate a "visual" meaning to such models can be even more challenging since they can also incorporate information from other fields (demographics, histopathology, genomics…). The human brain can only take into account a limited number of parameters in making a decision, therefore these multiparametric models will not be easily apprehended by end users. They will clearly demand a high level of precision and robustness in order to be accepted and relied upon to formulate a clinical decision. Within this context a rigorous process of model development (proper training) and validation (independent large cohorts) is needed, which is still far from being a standard, although some encouraging results have been published (16,17).

The current radiomics paradigm consists in adding quantitative information to the visual analysis by radiologists and nuclear medicine physicians, rather than replacing it entirely. For instance, it was recently shown that a set of semantic features obtained from visual assessment by radiologists could beneficially complement quantitative radiomics in determining EGFR mutations in lung cancer (18). However, a recent trend in medical imaging is to exploit techniques from the field of deep learning (19), with examples in image segmentation (20) or radiomics-type studies (21). This will further complicate the issue of association with visual analysis discussed above. Indeed, on the one hand the standard radiomics workflow relies on the extraction of carefully designed features based on domain expertise (e.g. a specific calculation in the intensity histogram or in a predesigned texture matrix), some of which are clearly inspired by the human visual system. On the other hand, deep learning methods automatically discover features from data and the representations useful for the task at hand using a general-purpose learning procedure such as convolutional neural networks (CNN). These require substantial amounts of data not easily available in the field of medical imaging, particularly in PET/CT. Potential solutions include transfer learning, consisting in using CNN trained for an unrelated task using large datasets, and adapt them to a different setting (22,23). If these tools were to replace advantageously the current workflow of radiomics, removing the need for tumor segmentation or the complex task of selecting relevant and reliable features (10,24), as well as an improved ability to handle standardisation issues (25), the relationship with visual analysis by experts would not simply be more difficult but certainly unnecessary to establish.

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441-6.

2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278:563-577.

3. Aerts H. Radiomics: there is more than meets the eye in medical imaging. In: *SPIE Medical Imaging 2016: Computer-Aided Diagnosis*.Vol 9785.; 2016:97850O-97850O-1.

4. Tixier F, Hatt M, Valla C, et al. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. *J Nucl Med*. 2014;55:1235-1241.

5. Kim D-H, Jung J-H, Son SH, et al. Quantification of Intratumoral Metabolic Macroheterogeneity on 18F-FDG PET/CT and Its Prognostic Significance in Pathologic N0 Squamous Cell Lung Carcinoma. *Clin Nucl Med*. 2016;41:e70-75.

6. Watabe T, Tatsumi M, Watabe H, et al. Intratumoral heterogeneity of F-18 FDG uptake differentiates between gastrointestinal stromal tumors and abdominal malignant lymphomas on PET/CT. *Ann Nucl Med*. 2012;26:222-227.

7. Hodgdon T, McInnes MDF, Schieda N, Flood TA, Lamb L, Thornhill RE. Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images? *Radiology*. 2015;276:787-796.

8. Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *J Magn Reson Imaging JMRI*. 2007;25:89-95.

9. 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.

10. Hatt M, Tixier F, Pierce L, Kinahan P, Cheze Le Rest C, Visvikis D. Characterization of PET images using texture analysis: the past, the present… any future? *European Journal of Nuclear Medicine and Molecular Imaging*. 2016:in press.

11. 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.

12. Desseroit M-C, D. Visvikis, Tixier F, et al. Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in Non-Small Cell Lung Cancer stage I-III. *European Journal of Nuclear Medicine and Molecular Imaging*. 2016:in press.

13. Desseroit M-C, Tixier F, Cheze Le Rest C, et al. Comparison of three quantization methods for the calculation of textural features in PET/CT images: impact on prognostic models in Non-Small Cell Lung Cancer. In: *IEEE NSS-MIC Conference Records*.; 2016.

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

15. Upadhaya T, Morvan Y, Stindel E, Le Reste P-J, Hatt M. A framework for multimodal imaging-based prognostic model building: Preliminary study on multimodal MRI in Glioblastoma Multiforme. *IRBM*. 2015;36:345-350.

16. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.

17. Desseroit M-C, Tixier F, Guillevin R, et al. Nomogram for NSCLC exploiting clinical staging, tumor volume and PET/CT heterogeneity features: development using support vector machines in a retrospective cohort and first validation results in prospectively recruited patients. *J Nucl Med*. 2016;57:437-437.

18. Velazquez ER, Liu Y, Parmar C, Narayan V, Gillies R, Aerts H. MO-DE-207B-08: Radiomic CT Features Complement Semantic Annotations to Predict EGFR Mutations in Lung Adenocarcinomas. *Med Phys*. 2016;43:3706-3706.

19. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.

20. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with Deep Neural Networks. *Med Image Anal*. 2016;35:18-31.

21. Ypsilantis P-P, Siddique M, Sohn H-M, et al. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. *PloS One*. 2015;10:e0137036.

22. Antropova N, Huynh B, Giger M. SU-D-207B-06: Predicting Breast Cancer Malignancy On DCE-MRI Data Using Pre-Trained Convolutional Neural Networks. *Med Phys*. 2016;43:3349-3350.

23. Huynh B, Drukker K, Giger M. MO-DE-207B-06: Computer-Aided Diagnosis of Breast Ultrasound Images Using Transfer Learning From Deep Convolutional Neural Networks. *Med Phys*. 2016;43:3705-3705.

24. Lian C, Ruan S, Denœux T, Jardin F, Vera P. Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction. *Med Image Anal*. 2016;32:257-268.

25. Lasnon C, Majdoub M, Lavigne B, et al. 18F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer. *Eur J Nucl Med Mol Imaging*. 2016:1-12.