



Images Are Data: Challenges and Opportunities in the Clinical Translation of Radiomics

Wei Mu¹, Matthew B. Schabath², and Robert J. Gillies³

Radiomics provides an opportunity to uncover image-based biomarkers through the conversion and analysis of standard-of-care medical images into high-dimensional mineable data. In the last decade, thousands of studies have been published on different clinical applications, novel analysis algorithms, and the stability and reproducibility of radiomics. Despite this, interstudy comparisons are challenging because there is not a generally accepted analytic and reporting standard. The ability to compare and combine results from multiple studies using interoperative platforms is an essential component on the path toward clinical application. The NCI supported study from van Griethuysen and colleagues published

in *Cancer Research* in 2017 proposed PyRadiomics: an open-source radiomics quantification platform for standardized image processing. Since released, it has become a frequently utilized analytic tool in the radiomics literature and has accelerated the capability of combining data from different studies. The subsequent challenge will be the design of multicenter trials with a fixed and immutable version of software, which is currently open-source, readily modified and freely distributed. Generally, this is accomplished with a commercial partner to navigate the regulatory processes.

See related article by van Griethuysen and colleagues, *Cancer Res* 2017;77:e104-7.

In the year 2000, the *New England Journal of Medicine* (1) identified cross-sectional imaging as one of the top 10 medical advancements in the last millennium. It is fitting that the quantitative analysis of these images with contemporary artificial intelligence (AI) methods, known as radiomics, has matured to the cusp of clinical adoption here in the 50th year since passage of the National Cancer Act of 1971 (2).

Medical imaging is a necessary examination technology in oncology care. With the technological improvement in scanning techniques and AI algorithms, information from these images can be swiftly converted from qualitative assessment to quantitative and structured data. In radiomics, medical images are converted into high-dimensional mineable data for the subsequent analysis with AI to uncover imaging-based biomarkers for diagnosis, prognosis, and to predict treatment response. Because the concept of radiomics was first introduced in 2012 (3), thousands of radiomics studies on numerous pathologies using different algorithms have been published, sometimes with spectacular performance results. For example, recent early detection studies led by Google Inc. on lung cancer using CT (4) and breast cancer using mammography (5) have clearly shown the potential of AI in medical image analysis. However, in general, these studies have lacked rigor and are difficult to reproduce, as most analytic approaches are bespoke for particular applications. Even though deep learning networks and associated weightings should be easily distributed, the aforementioned studies failed to disclose the code used for training the

models, which presents an obstacle to reproducibility. There is an increasing recognition in AI-based research, including radiomics, that data reporting should be Findable, Accessible, Interoperable, and Reusable (FAIR; ref. 6). This focus and standardized data collection has made both the tools created and the data captured readily shareable across multiple networks.

In conventional radiomics (in contrast to that which is deeply learned), hand-crafted features are calculated from user-defined volumes of interest (VOI), such as lung nodules detected in low-dose CT images. This is comprised of volume delineation to accurately define the entire VOI, feature extraction where individual radiomics features are calculated from the VOI, feature selection to reduce the number of radiomics by removing correlated, unstable, and nonreproducible features, and then task-oriented predictive model training, validation, and testing to identify the most informative features or algorithm (Fig. 1). In this manner, the radiomics features are usually extracted with in-house software, which hinders the reproducibility and comparison of different results due to differences in image preprocessing, feature nomenclature, mathematical definition, software implementations, and algorithms (7). To address this, the image biomarker standardization initiative (IBSI) emerged to provide standardized radiomics feature nomenclatures, definitions, formulae, image processing workflow, a radiomics software verification tool, and reporting guidelines (8). Compliant to IBSI, in a landmark publication in *Cancer Research* in 2017 (9), van Griethuysen and colleagues developed PyRadiomics, a comprehensive open-source platform, which helped establish a reference standard for radiomics analyses.

The PyRadiomics platform can extract all the commonly used and IBSI compliant handcrafted features including first order statistical features, texture features from original or filtered images, as well as shape descriptors, and supports two-dimensional and three-dimensional (3D) segmentations of different imaging modalities (such as CT, PET, and MRI). Developed with Python, PyRadiomics can be installed on any system and can be used as a standalone or in combination with an open-source image analytics and display tool such as 3D Slicer. Owing to the flexibility, less dependence on programming skills, and minimum risk of programming errors,

¹School of Engineering Medicine, Beihang University, Beijing, P.R. China.

²Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida. ³Department of Cancer Physiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida.

Corresponding Author: Robert J. Gillies, Department of Cancer Physiology, Moffitt Cancer Center, 12901 Magnolia Dr, Tampa, FL 33612. Phone: 813-745-8355; E-mail: robert.gillies@moffitt.org

Cancer Res 2022;82:2066-8

doi: 10.1158/0008-5472.CAN-22-1183

©2022 American Association for Cancer Research

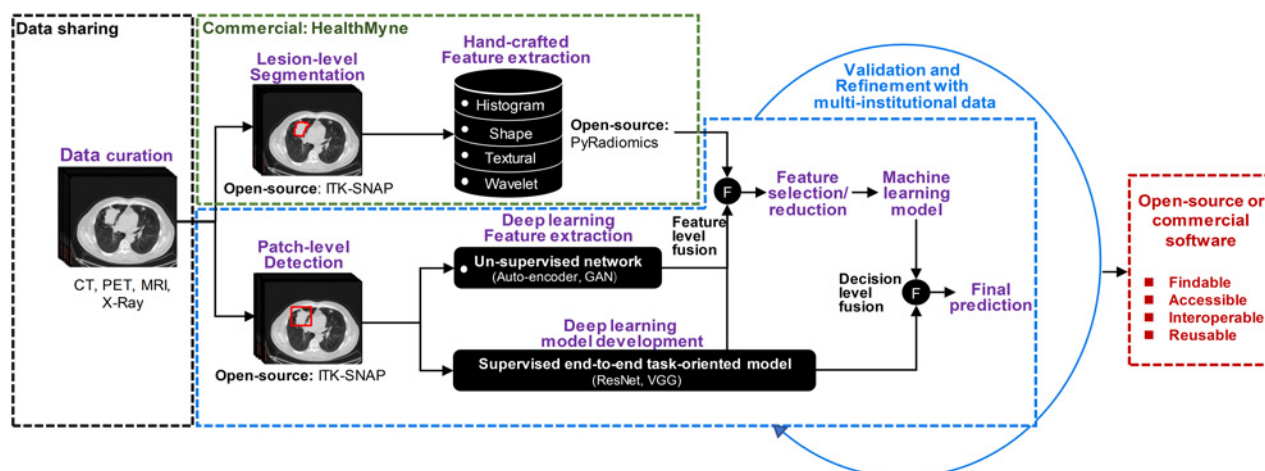


Figure 1. Image-based analytic pipelines and qualification for radiomics biomarkers. Conventional handcrafted radiomics features are calculated from user-defined VOI through a series of steps (black and green boxes) that include volume delineation, feature extraction/calculation, feature selection/reduction, and then task-oriented predictive model development and validation (right end of the blue box). Deep learning (blue box) methods that do not depend on volume delineation of a VOI and radiomics features are not extracted. Rather, a patch (e.g., bounding box), an entire image (i.e., a single slice), or an entire volumetric image series can be an input into a deep learning network for the task-oriented predictive model development and validation. To ensure the rigorous management and stewardship of data from image analyses, a radiomics pipeline should be standardized, open-source, and aligned with FAIR standards (red box).

PyRadiomics has become a frequently reported analysis tool in radiomics research since it was introduced. As a consequence, radiomics studies have become more reproducible and comparable.

Though the number and calculation method of features are fixed in PyRadiomics, several efforts have validated its potential to mine more information from images to improve the performance of radiomics models. In oncology studies, due to tumor heterogeneity, radiomic-identified peritumoral and intratumoral regions (habitats) each have important metabolic characteristics that impact response to therapy. Furthermore, changes in radiomics features from dynamic follow-up images (delta-radiomics) have also been quantified to discover additional diagnostic and prognostic information. As PyRadiomics is an open-source platform and has a high degree of flexibility, characterizing the different user-defined or automatic generated regions of interest helps to deepen the diversity of radiomics features while maintaining reproducibility and comparability.

After the release of PyRadiomics, more IBSI-compliant software, such as LIFEx, CERR, and MITK Phenotyping have been developed with other languages including Java, MATLAB, and C++. Most of these software options are open source with well-documented mathematical equations. Meanwhile, several commercial software platforms represented by HealthMyne or MIM, for example, have been developed to process and extract radiomics features. Commercial software platforms have an important role in helping to standardize radiomics analyses by providing stable results, being friendly to new researchers, and accompanied by technical support. However, commercial software, different from the open-source software, protects the source code and could only be used as a “black box,” which limits the flexibility in expanding the existing features.

With the fast development of deep learning algorithms, deep learning has been introduced in radiomics. Current deep learning-based radiomics analyses can be categorized as two different types: supervised end-to-end discriminative deep learning models developed to establish the relationship between medical images and clinical

outcome, and unsupervised generative deep learning models developed to learn rich features that describe the data for the consequential prediction tasks as feature-engineered radiomics analysis. Owing to the self-learning characteristics of deep learning, the deep learning models can mine more task/data-related information to increase the prediction accuracy and decrease the dependence on accurate delineation. Therefore, deep learning-based radiomics has quickly become another mainstream direction for the field during the last several years. In contrast to the handcrafted feature-engineered radiomics analyses, utilization of deep learning models should yield more reproducible results without the effect of feature nomenclature, mathematical definitions, or software implementations once the models have been well trained. However, these deeply learned models have a higher requirement for diverse and large training sets, compared with conventional radiomics, often requiring multi-institutional studies. Open-source code for each proposed model is of great importance in developing new, bespoke applications, yet there are significant challenges to implement this in a standardized fashion in a commercialized setting across multiple sites.

While radiomics is still at an early stage of development, it is too early to conclude which software or pipeline is optimal for each task. Hybrid or ensemble methods that combine handcrafted radiomics features and deep learning by using decision-level and feature-level fusion are increasingly being utilized in more recent studies. In one approach for decision-level fusion, the feature-engineered and deep learning models are trained separately and their outputs combined with different voting strategies to achieve the final decision (10). The feature-level fusion trains a classifier with the handcrafted features and uses deep learning features for the final prediction. In this way, both reference standard and open-source are needed. Overall, no matter which approach is used, harmonization of certain settings is needed to guarantee reproducibility and comparability.

In a single package, PyRadiomics addresses many challenges in image analysis including ease of use, flexibility, reproducibility, standardization, transparency, and capability for expansion and

customizations. PyRadiomics, which has seen tremendous uptake and utilization since its release, has revolutionized image analysis and radiomics research by providing a reference standard that is readily accessible and open-source. The next step will be multi-center testing and deployment of use-specific locked down versions, which could make radiomics data reimbursable for decision support.

References

1. Looking back on the millennium in medicine. *N Engl J Med* 2000;342:42–9.
2. The national cancer act of 1971. *J Natl Cancer Inst* 1972;48:577–84.
3. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6.
4. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25:1319.
5. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94.
6. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. Addendum: The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2019;6:6.
7. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749–62.
8. 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;295:328–38.
9. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104–7.
10. 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.

Authors' Disclosures

R.J. Gillies reports non-financial support from HealthMyne, Inc. during the conduct of the study and other support from HealthMyne, Inc. outside the submitted work; in addition, R.J. Gillies has a patent for 60/865,544 issued. No disclosures were reported by the other authors.

Received April 8, 2022; accepted April 8, 2022; published first June 6, 2022.