Radiology

How Can Radiomics Be Consistently Applied across Imagers and Institutions?

Peter Steiger, PhD • *Rohit Sood, MD, PhD*

From the Department of Scientific and Medical Services, Parexel Informatics, Parexel International, 2 Federal St, Billerica, MA 01820. Received January 7, 2019; revision requested January 8; revision received January 10; accepted January 10. **Address correspondence to** P.S. (e-mail: *peter.steiger@parexel.com*).

Conflicts of interest are listed at the end of this article.

See also the article by Orlhac et al in this issue.

Radiology 2019; 291:60–61 • **https://doi.org/10.1148/radiol.2019190051** • **Content code: IN** •

The advent of CT and MRI made the concept of image-based quantitation feasible. Early efforts involved drawing rulers, regions of interest, and volumes of interest, in addition to evaluating dimensions, average densities, and other imaging characteristics. This added a quantitative ability to a radiologist's otherwise purely qualitative assessment of images. It exploited the capability of computers to do something that human observers are not good at: making measurements and quantitative assessments. It opened up the ability for radiologists to arrive at conclusions that went beyond what could be derived by interpretation of the anatomy alone.

After the initial enthusiasm about quantitative imaging, it became apparent that not all quantitative information was reliable: CT scanners had different calibrations, beam-hardening effects created field of view inhomogeneities, and MR imagers had geometric distortions. The complex physics of ultrasonic imaging brought additional challenges. Over the years, many methods were developed for a variety of quantitative assessments, such as bone density, muscle volume, cartilage volume, and composition, as well as liver fat and iron content. Texture-based quantitative image analyses also show promise for multiple assessments, including quantifying trabecular structure, evaluating fibrotic lung disease, differentiating Gleason scores in prostate cancer MRI, and assessing primary colorectal tumor heterogeneity. Despite the vast amount of research, relatively few quantitative imaging methods are used widely in clinical settings, such as quantitative CT for bone densitometry or echocardiography. Some applications resulted in dedicated medical devices, such as dual x-ray absorptiometry for the management of osteoporosis and computer-assisted diagnosis for detecting breast cancer on mammograms and lung nodules on chest radiographs.

The introduction of radiomics has brought with it the vast expansion of the promise of quantitative and objective assessment of images. Interpretations are no longer limited to features like area, volume, and histogram-derived metrics; they can include hundreds of different features including shape, gray-level run-length matrices, Haralick texture, heterogeneity, coarseness, or busyness (1). Putting such higher dimension image characteristics into the context of increasingly accessible clinical information about patients holds promise for evidence-based clinical decision support (1).

With the expansion of knowledge on how to apply radiomics also comes the recognition that variation in image acquisition parameters and equipment can have a big effect on outcome, just like in quantitative imaging. Methods have been proposed to deal with this challenge, but most of these methods involve the laborious resampling of images and extensive characterization of the image generating equipment to be effective.

In this issue of Radiology, Orlhac et al (2) adapt a method originally used in genomics to correct variations in radiomic measurements caused by different imagers and imaging protocols (2). The proposed method is based on a statistical method called ComBat, which is readily available in the open-source R statistical programming language (R Foundation for Statistical Computing, Vienna, Austria). Unlike other previously published methods, this approach does not require images to be modified. It allows for correction of radiomic measurements on the basis of their distribution and knowledge of covariates. The authors tested their method with one publicly available phantom data set and two patient data sets from patients with lung cancer. They convincingly showed that their method reduced imager-induced variability without sacrificing diagnostic sensitivity. Their article explains the method clearly and provides all the references needed to replicate the work. This should encourage others to apply this method and test it in other radiomics studies and applications.

Clinical trials in support of drug approval go to great lengths to control the variation in image acquisition by carefully specifying and then monitoring image acquisition parameters and image quality. Best practices have been published by the U.S. Food and Drug Administration (3). To our knowledge, no other regulatory agency has published similar guidelines and therefore this guidance has global applicability. However, it is not always possible to control image quality in this way. Imaging-based biomarkers have an increasingly important role in getting new drugs to market, not just in oncology. With the growing availability of real-world data, situations arise in which research questions need to be evaluated on the basis of data acquired in less controlled environments. Methods like those presented in Orlhac et al (2) can be powerful tools to evaluate these data.

Studies have already demonstrated the ability of radiomics to estimate the Gleason grade for prostate cancer (4), predict survival in glioblastoma (5), or to be used as prognostic indicators for patients with early stage lung cancer (6), just to name a few. Radiomics holds the promise to become a tool at the disposal of the radiologist to expand the qualitative interpretation

of the image, with additional quantitative information that can provide functional and prospective information not evident from the image alone. More studies are needed to fulfill this promise. The proposed algorithm has been shown to be effective in both thin- and thick-section CT images. However, thin-section CT images reconstructed by using high-spatial-frequency algorithms or thin-section CT are more useful for detecting the higher order image features, which is at the heart of radiomics. For example, thin-section CT of the lung has been successfully used for quantitative assessment of lung fibrosis by using data-driven texture analysis in patients diagnosed with idiopathic pulmonary fibrosis (7), a form of interstitial lung disease. It is important to note that standard-of-care CT protocols in most lung cancer studies specify 5-mm-thick sections, thus limiting the application of radiomics in these types of studies. Therefore, the general success of radiomics in lung cancer and oncology will in part depend on the development and adoption of tailored image acquisition techniques for quantitative feature analysis. Radiomics will benefit from an extension of efforts already underway to standardize quantitative imaging, spearheaded by the Quantitative Imaging Biomarkers Alliance (8).

The study by Orlhac et al (2) has limitations. The compensation method did not work as well on images in patients as it did on the phantom images. This was particularly evident in one of the patient data sets, in which significant differences remained in 30% of the patients, twice the rate of the second data set. Still, although the same level of significance was not reached, the authors were able to demonstrate a meaningful improvement. Further, only patients with lung cancer were included in this study. This method will need to be tested with other disease states, including nononcology studies, where radiomics may be applicable.

Substantial hurdles remain until radiomics can become a routine tool in the radiology reading room of the future, as eloquently explained by Gillies et al (1). Among them is the need to validate any radiomics biomarkers in prospective multicenter studies. The variability introduced by the wide variety of available equipment and imaging protocols must be controlled to allow these radiomic biomarkers to be used in a broader manner. The method presented by Orlhac et al (2) may have an important role in this research.

Disclosures of Conflicts of Interest: P.S. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed money paid to the author from Parexel for employment. Other relationships: disclosed no relevant relationships. R.S. disclosed no relevant relationships.

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–577 .
2. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. Radiology 2019;291:53–59.
3. U.S. Department of Health and Human Services, Food and Drug Administration. Clinical Trial Imaging Endpoint Process Standards. Washington, DC: Federal Register, 2018.
4. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A 2015;112(46):E6265–E6273 .
5. Diehn M, Nardini C, Wang DS, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A 2008;105(13):5213–5218 .
6. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5(1):4006 .
7. Humphries SM, Swigris JJ, Brown KK, et al. Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. Eur Respir J 2018;52(3):1801384 .
8. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC; Group. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. Radiology 2011;258(3):906–914 .