How Far Are We from Using Radiomics Assessment of Gliomas in Clinical Practice?

Rajan Jain, MD • Yvonne W. Lui, MD

From the Department of Radiology and Neurosurgery, NYU School of Medicine, 660 First Ave, 2nd Floor, New York, NY 10016. Received August 30, 2018; revision requested September 5; final revision received September 5; accepted September 7. Address correspondence to R.J. (e-mail: rajan.jain@nyumc.org).

Identifying important characteristics from an image was described for aerial photographs as early as 1955 and eventually by Haralick et al in 1973 using computable texture features (1). Radiomics is a more recent and fancy name given to this field of study in which high-throughput data are extracted and large amounts of quantitative imaging features are generated from medical images using data-characterization algorithms and computers. In a way, it can be thought of as reverse engineering of medical images—for decades it has been the diagnostic imaging unit manufacturers’ aim to create images from data acquired from human tissue, in some cases postprocessing those data to make the images “prettier” to the viewing eye such as through the use of smoothing algorithms or improving contrast-to-noise ratio while in the process altering, hiding, or potentially losing acquired information. By reversing that process, radiomics seeks not only to go back to the vast data used to create images in the first place but also to uncover the patterns of imaging phenotypes hidden within those data that could be clinically useful.

In brief, image acquisition, segmentation, feature extraction, and feature selection are some of the essential steps involved in radiomics analysis. Thus far, in tumor radiomics, commonly extracted features generally fall into four categories: first-order statistics, texture features, wavelet features (features from a transformed space), and shape features. First-order statistics derive features such as mean, variance, kurtosis, skewness, and entropy that describe the histogram distribution of the entire tumor. Texture features are called second-order statistics because they capture spatial mutual dependencies of the image voxels such as homogeneity, contrast, gray-level nonuniformity, cluster tendency, and harder-to-picture features such as short run emphasis. Two methods commonly used to derive texture features include gray-level co-occurrence matrices and gray-level run-length matrices. To complicate matters, images can be mathematically transformed to then extract features in the transform space, such as wavelet transformation yielding features dependent on spatial frequencies. Finally, shape features such as volume, surface area, sphericity, compactness, and flatness may be extracted on the basis of the shape of the tumor border.

The first three of these subgroups of features are also known as “agnostic” features; these are mathematically extracted quantitative descriptors and historically have not been part of the typical radiologist’s lexicon (2). On the other hand, shape features are often referred to as “semantic” features precisely because some of these have been intrinsic to the radiologist’s lexicon for years (2), although radiomics additionally includes quantification of these features with computer assistance. Because radiology training and clinical practice focus on image pattern recognition from “processed and presented” medical images rather than tasks of quantitative data analysis, radiologists are not familiar with, let alone accustomed to, using the majority of agnostic features and terminology. Hence, to think that such features will soon be incorporated into routine radiology reports may be premature.

In this issue of *Radiology*, Bae et al (3) extracted 796 radiomic features (702 texture features, 70 shape features, and 24 apparent diffusion coefficient [ADC] histogram features) from multiparametric MRI and used machine learning to demonstrate the added value of radiomics analysis to clinical and genetic features for survival prediction in 217 patients with glioblastoma. The authors conducted random survival forest (RSF) analysis and exploited open-source packages to build the pipeline from image processing to machine learning by using separate training (163 patients) and test (54 patients) sets to validate their results. While the benefit in area under the receiver operating characteristic curve overall was somewhat incremental in this study, the results show the potential influence that radiomic features may have in the future of image and lesion analysis.

Having such a plethora of numbers and types of imaging features is exhausting to even think about. Many are abstract and hard to picture. There is the concern of overfitting a model and not learning the true basis of a decision. One way to help address the problem of overfitting is first to perform feature selection, choosing the most powerful features and reducing the number of features inputted into the model. Bae et al used variable-hunting feature selection to whittle their original 796 features down to the 18 most useful ones (two first-order, nine texture, and seven shape features).

In addition to feature selection, model choice is another variable that further complicates an analysis such as this. Here, the authors chose an RSF model, using standard 10-fold cross validation on the training set and validated on the test set. Compared with commonly used Cox regression, RSF has two main reported advantages: (a) RSFs are free from proportional hazard assumption and are fully nonparametric (thus, the prognostic value of the RSF model is not limited, even when some features of the model are time-dependent [eg, risk associated with the feature changes...
Will We Soon Use Radiomics Assessment of Gliomas in Clinical Practice?

Will We Soon Use Radiomics Assessment of Gliomas in Clinical Practice?

over time] or have interaction with other features [4]); and (b) RSF models are more robust to noise variables than Cox regression, yielding stable performance (4). These strengths are especially advantageous for high-dimensional data such as radiomics, where not all features can be strictly and individually controlled. Using this method, Bae et al identified their 18 radiomics features and were able to stratify patients into low- and high-risk groups in both training and validation sets, showing slight improvement in overall and progression-free survival prediction.

Studies such as this one raise several important questions: How do these imaging features relate with what we already know about disease biology? Can such additional information add to our understanding of a disease (eg, gliomas)? Does this information change or enhance what radiologists already know about imaging–tumor biology association? Some work has begun in this arena: First-order features obtained from histogram-based methods have been shown to relate to tumor cellularity (5). Textural features, on the other hand, reflect tumor heterogeneity. They are potential markers for tumor aggressiveness and, possibly, response to therapy (2,6). Tumor shape features (the only “semantic” features) have also been shown to relate to tumor aggressiveness (2). Despite growing knowledge about radiomic features, the difficulty that remains is how you and I might incorporate results such as these into clinical practice.

Another issue is that of “abundance”—too much data and too many features, not only to understand but also to practically incorporate into standard radiology dictations and reports. The authors summarize a number of other articles in the literature that attempt to predict outcome in glioblastomas using similar but different techniques, finding different features and having variable results. It remains unknown which features among the myriad available are truly most important. In current practice, radiology reports include no more than three or four semantic features to describe a tumor (eg, a tumor with irregular borders). This is traditionally how clinical radiologists relate imaging findings to tumor biology—on the basis of a shared knowledge gained from radiologic-pathologic associations and, more recently, from genomic associations (also known as radiogenomics or imaging-genomics).

Last, like other quantitative data, radiomic features are sensitive to the methods and software used in the analysis pipeline, potentially limiting the reproducibility and generalizability of any given radiomic models. Sharing code and open-sourcing pipelines, as done by the authors in this study, is a critical way to help address this. Nevertheless, external validation is still mandatory to test the accuracy of radiomics models across different populations and in larger studies. Another issue is how and whether to “trust” information generated by an algorithm, which seems very much like a black box. This is an important topic beyond the scope of this editorial, and the readers are referred to Ribeiro et al (7); giving meaning to the black box can also form the basis for future work in medical imaging, radiomics, and machine learning.

In conclusion, radiomics appears to be a powerful potential tool in the armamentarium of brain tumor imaging assessment. The work of Bae et al suggests that radiomics can augment survival prediction in glioblastomas, although the added benefit at this point is slight. With the continued growth of computational power, radiomics is here to stay; however, there exist several challenges before radiomics can reliably be integrated into mainstream clinical practice.

Disclosures of Conflicts of Interest: R.J. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Cancer Panels; receives royalties from Thieme. Other relationships: disclosed no relevant relationships. Y.W.L disclosed no relevant relationships.

References

2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–577.