Radiology

REVIEWS AND COMMENTARY • EDITORIAL

Data Partitioning and Statistical Considerations for Association of Radiomic Features to Biological Underpinnings: What Is Needed

Michael A. Jacobs, PhD

Dr Jacobs is an endowed professor of radiology and vice chair of research at the Department of Diagnostic and Interventional Imaging, McGovern Medical School at the University of Texas Health Science Center. His current research interests include developing radiologic methods for detection, monitoring, and treatment of many diseases. He has been trained in diagnostic medical physics, advanced mathematics, and engineering. His recent work has been in radiomics, whole-body MRI, and computer science to aid in diagnosing and detecting different types of disease burden.

Over the past decade, radiomics, or texture analysis, has been increasingly investigated for its utility as a potential biomarker derived from different radiologic images (1,2). There are basically two types of radiomic features: first order (statistical features) and second order (gray-level matrix [fine and course features]). They are generated as output as either single radiomic features (3–5) or multiparametric radiomics features (6). Most studies use single handcrafted regions of interest (3–5) or full radiomic images based on the tissue of interest (6). Several studies have shown some correlation of radiomic features with other important clinical parameters, which one day could make radiomics a standard-of-care parameter used in a clinical setting. However, radiomics is still an investigative tool, with several groups actively pursuing standardization methods for more accurate radiomic features (7,8).

In this issue of *Radiology*, the article by Gidwani et al (9) brings into focus the careful considerations of data partitioning and statistical methods that are needed to ensure reproducible data analysis without resulting in "inflated" measures of accuracy and spurious associations when using radiomics coupled with machine learning (ML) methods. Moreover, a recent publication in *Radiology* has highlighted these types of concerns as well (10). These reports are very timely and needed to further progress the interpretation of radiomic features as they are related to biology for a more accurate prediction of potential clinical association or significance.

The authors (9) report and demonstrate that incorrect data partitioning can lead to a very considerable boost, at least 1.4-fold, in the performance of the radiomic features when using ML to obtain the most significant factors based on the area under the receiver operating characteristic curve (AUC) and correlation analysis in overall survival. The radiomic features were derived from two public data sets consisting of low-grade gliomas, head and neck cancer, and further testing of radiomic features with association of gene array scores. The findings reported could have implications for identifying which intrinsic features are important, and the authors provide a roadmap to strengthen the testing of radiomic-ML pipelines. For example, using a model of data leakage in their simulated radiomic feature set with the different ML models resulted in high correlations and AUC metrics. When the data leakage was "corrected" in the data set, the results become inconclusive with nondiagnostic AUC values. Another major implication of the results of this study is that Gidwani et al (9) describe significant correlations between radiomics and gene array data by using simulated radiomic features that had no biologic meaning. This is clearly demonstrated when mixing high-dimensional data sets, which can be problematic due to the sparsity of the data points and can lead to the spurious correlations. As noted in the article, care needs to be taken to ensure that no data leakage can occur and to consider if the results make practical sense.

The authors present a clear direction on how to avoid these potential pitfalls when using the radiomic-ML pipeline through a series of questions investigators may ask while designing a study. These are summarized as follows: *(a)* Is a sample size estimate performed to determine the significance of the result? *(b)* Is partitioning applied correctly, and is it consistently observed through the different steps of ML application? *(c)* Have reproducibility and multiple hypotheses and correction methods (if applicable) been applied; and *(d)* Is there an external data set available for testing the model? Also, some investigators need to test if the radiomic results and correlations make sense with any quantitative imaging metrics (eg, T1, T2, or apparent

From the Department of Diagnostic and Interventional Imaging, McGovern Medical School at The University of Texas Health Science Center at Houston (UTHealth Houston), 6431 Fannin St, Room R172, Houston, TX 77030; and The Russell H. Morgan Department of Radiology and Radiological Science and Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University, Baltimore, Md. Received November 21, 2022; revision requested November 22; revision received November 26; accepted November 30. **Address correspondence to** the author (email: *Michael.A.Jacobs@uth.tmc.edu*).

M.A.J. is supported by National Institute of Health (NIH)/The National Cancer Institute (NCI) (grant U01CA140204 [Renewal]), NIH/NCI (grant P30CA006973), Defense Advanced Research Projects Agency (DARPA) (grant DARPA-PA-20-02-11-ShELL-FP-007), NIH/National Institute of Diabetes and Digestive and Kidney Diseases (grant U01DK127400), and NIH/National Heart, Lung, and Blood Institute (grant 1R01HL149742).

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

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

diffusion coefficient of water mapping [MRI], standard uptake values [PET], or Hounsfield units [CT]). Finally, the model design introduced by Gidwani et al (9) may be able to reduce bias and spurious correlations in radiomic research and help by giving more insight and reliability to the radiomics-ML pipeline when applied to radiologic imaging or data sets.

Disclosures of conflicts of interest: M.A.J. Member of *Radiology* editorial board; US patent 8,380,281 for License Diagnsoft; patents planned, issued, or pending: US patent 8,380,286, US patent 8,380,281, US Patent 9,008,462, US Patent 9,256,966, US Patent 20,160,132,754, US Patent 10,388,017 B2, US Patent 11324469 B2, US Patent 20,180,189,384, US Patent Application No. 63/178,705. (WIPO) World Intellectual Property Organization for Trade Patents: WO2013177586, WO2015017632, WO2015164517.Melt Curve Classifier for Reliable Large-scale Genotyping of Sequence Variants, Pending/Filed: JHU Reference Number #C12028, JHU Reference Number #D13500, JHU Reference Number #D13769, JHU Reference Number #D14297, JHU Reference Number #D15143, JHU Reference Number #C16639; editor of *Journal of Biomedicine and Biotechnology: Radiology BioMed Research International*, *Expert Review of Precision Medicine and Drug Development*, *Radiology*, and *Medical Physics*.

References

1. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. Expert Rev Precis Med Drug Dev 2016;1(2):207–226.
2. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–762.
3. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30(9):1234–1248.
4. 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(4):441–446.
5. Kontos D, Winham SJ, Oustimov A, et al. Radiomic Phenotypes of Mammographic Parenchymal Complexity: Toward Augmenting Breast Density in Breast Cancer Risk Assessment. Radiology 2019;290(1):41–49.
6. Parekh VS, Jacobs MA. Multiparametric radiomics methods for breast cancer tissue characterization using radiological imaging. Breast Cancer Res Treat 2020;180(2):407–421.
7. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295(2):328–338.
8. Soliman MAS, Kelahan LC, Magnetta M, et al. A Framework for Harmonization of Radiomics Data for Multicenter Studies and Clinical Trials. JCO Clin Cancer Inform 2022;6(6):e2200023.
9. Gidwani M, Chang K, Patel JB, et al. Inconsistent partitioning and unproductive feature associations yield idealized radiomic models. Radiology 2023;307(1):e220715.
10. Moskowitz CS, Welch ML, Jacobs MA, Kurland BF, Simpson AL. Radiomic Analysis: Study Design, Statistical Analysis, and Other Bias Mitigation Strategies. Radiology 2022;304(2):265–273.