From Code to Bedside: Introducing Predictive Intelligence to Interventional Oncology

Julius Chapiro, MD, PhD • James S. Duncan, PhD

Julius Chapiro, MD, PhD, is a junior principal investigator who leads an interdisciplinary team of basic, translational, and clinical scientists within the Interventional Oncology Research Laboratory at the Yale Department of Radiology & Biomedical Imaging. His federally, foundationally, and industry-sponsored research focuses on developing new quantitative imaging biomarkers, predictive instruments, and imaging techniques for the diagnosis, characterization, and image-guided therapy of hepatocellular carcinoma.

James S. Duncan, PhD, is the Ebenezer K. Hunt professor of biomedical engineering, associate chair and director of undergraduate studies in the department of biomedical engineering as well as the vice chair for bioimaging sciences research in radiology. With his internationally renowned, NIH-funded, and award-winning research, he pioneered the development of model-based mathematical strategies for the analysis of biomedical images, applied to complex neurologic conditions, cardiac disease, and cancer.

Liver cancer is a growing global health care challenge and represents the third most common cause of cancer-related deaths worldwide. The majority of all patients with both primary and metastatic liver cancer are ineligible for curative surgical therapies at diagnosis. A large subset of these patients, especially those with hepatocellular carcinoma (HCC) which is diagnosed at intermediate to advanced stages, will undergo minimally invasive image-guided therapies such as transcatheter arterial chemoembolization (TACE) (1). Such treatments continue to be the mainstay guideline-approved therapeutic options for this group of patients, despite recent advances in systemic therapies that primarily involve novel immune-checkpoint inhibitors and other molecular-targeted agents. Therapeutic decisions for these patients are usually being made by consensus panels of interdisciplinary physician teams that are assembled in tumor boards. After thorough review of imaging, pathologic, laboratory, and clinical data, tumor boards determine the presumably best course of action based on a combination of individual physician expertise and evidence-based recommendations that are ideally derived from prospective randomized clinical trial data. The recent decade has witnessed a true explosion of data, both in terms of published information as well as directly patient-derived information that stems from novel multiparametric imaging, laboratory, genomic, and clinical pa-

rameters. In a scenario in which an overwhelming amount of new data is added to our knowledge base faster than it can be effectively processed and translated into meaningful therapy guidelines, it is almost certain that patients no longer benefit from the most recently added evidence as well as they should be. The lack of quantitative outputs from the underlying data may additionally result in highly variable therapeutic algorithms for disease management. In fact, tumor board decision can be overwhelmed by data chaos and conflicting subjective opinions. Recent data indicated that only a minority of centers strictly adhere to guidelines and recommendations such as the Barcelona Clinic Liver Cancer (BCLC) staging system (2). This is especially prevalent for patients with intermediate to advanced disease stages (BCLC stage B and C), a group of patients with vastly heterogeneous characteristics and presentations. All of the aforementioned circumstances clearly underline the urgent and unmet need for a novel approach to clinical and imaging information that would help us issue therapeutic recommendations that will translate quantitative data into better outcomes.

In their most recent article, Morshid et al propose a mechanism to address an aspect of the aforementioned global dilemma by using machine learning to extract quantitative parameters from baseline CT imaging in patients undergoing TACE with the ultimate goal to predict response after therapy (3). The authors apply a robust and familiar machine learning model, a random forest classifier, to combine BCLC stage and quantitative imaging features as inputs for the decision tree and investigate their combined potential to predict outcome. Their retrospectively collected dataset included a total of 105 patients with HCC with single tumors and baseline multiphase CT, all of which were treated with TACE as the first-line or initial bridging therapy. The authors defined time to progression (TTP), based on modified Response Evaluation Criteria in Solid Tumors, as their primary outcome marker. Lesions were divided into TACE-susceptible and TACE-refractory tumors based on a 14-week cutoff value of TTP. Two convolutional neural networks (CNNs), one for the tumor and the other for background liver segmentations, respectively, were then constructed and used for automated segmentation and for the extraction of specific imaging features. The CNNs were trained on two separate datasets, consisting of 130 CTs from The Medical Image Computing and Computer Assisted Intervention (MICCAI) Liver Tumor Segmentation (LiTS) challenge and 105 multiphase CT images of the study cohort. At first,

From the Yale Interventional Oncology Research Laboratory, Department of Radiology and Biomedical Imaging, Yale University School of Medicine, 300 Cedar St, TAC-N312A, New Haven, CT 06510. Received August 6, 2019; accepted August 12. Address correspondence to J.C. (e-mail: julius.chapiro@yale.edu).

See also the article by Morshid et al in this issue. Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2019; 1(5):e190139 • <https://doi.org/10.1148/ryai.2019190139> • Content codes: GI IN IR • ©RSNA, 2019

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

all regions of interest were manually segmented by expert radiologists. The manual segmentations of liver tissue and tumors were then considered ground truth. The authors validated the accuracy of automated segmentations using the Dice similarity coefficient to compare the machine learning-generated segmentations with the manual ones. An open-source Python package, Pyradiomics, which includes a library with a large set of prefabricated features, was used for feature extraction. The authors then applied the Boruta feature selection method to identify features that were most predictive of response to TACE. Among others, the search identified overall tumor volume and maximum two-dimensional axial diameter of the background liver as the most predictive parameters for outcome. Subsequently, the authors applied the random forest classifier for binary classification as the target output. First, a random forest classifier was trained with the BCLC stage as the only input. The observed accuracy was 62.9% (95% confidence interval [CI]: 52%, 72%), and increased to 74.2% (95% CI: 64%, 82%) after adding the predictive imaging features extracted from the automated segmentations. Then, a second random forest classifier was trained based on features extracted from manual segmentations. Interestingly, this experiment demonstrated a lower overall accuracy (67.6% [95% CI: 57%, 76%]). The authors explained this finding with the presumably higher subjective variability in the manual segmentations. Overall, the authors concluded in their discussion that adding specific segmentation-based imaging features to the well-known BCLC staging system substantially improved the ability to predict outcome compared with the BCLC staging system alone, especially in combination with automatically extracted features.

The approach to patient data, as well as the experimental design of this study overall, stands on solid ground. The authors validate prior works in the field and provide additional proof of principle for a data-driven decision support system in a clinical tumor board scenario (4). An important take-home message of this research is that tapping into previously unknown imaging characteristics that are potentially invisible to the “naked” radiologic eye may unleash the value of hidden imaging data points in the process of outcome prediction and treatment allocation. Although all the proposed conclusions are valid, it is also important to keep in mind that retrospectively collected data adds significant limitations that may affect the overall applicability of the proposed approach. As such, the authors used a relatively heterogeneous cohort that consisted of patients treated with either drug-eluting bead TACE or conventional oil-based TACE, which may have introduced inconsistencies with respect to possible differences in outcome. The proposed combination of imaging and clinical data resulted in a valuable, yet only incremental improvement of the predictive value, reaching an increase in accuracy of just above 10% as compared with the well-known staging system taken alone. While certainly limited by availability of larger cohorts of patients with well-annotated

data, interventional oncology applications should follow in the footsteps of others in related fields that provide a significantly higher added value with similar deep learning-based approaches (5,6). An important strength of the study is the fact that the authors used a publicly available dataset from the MICCAI LiTS challenge for external validation of their CNN. This underpins the need for and added value of such public libraries and encourages us to expand such data collections to help investigators design data-driven solutions, such as the one proposed by Morshid et al. The translation of the proposed approach from code to bedside will ultimately require an organized interdisciplinary effort that will allow for such algorithms to be trained on substantially larger collections of well-annotated data with the appropriate clinical questions asked. The fundamental ability to predict therapeutic outcomes prior to allocating patients to a specific therapy arm with greater accuracy than what is currently available represents a major health care challenge, both scientifically and economically. It will be thus up to the community to expand upon the presented research and map out a specific translational pipeline for such decision support tools with the ultimate goal to introduce predictive intelligence in interventional oncology for the benefit of clinical patient care.

Acknowledgments: We thank Alexandra Petukhova and Ahmet Küçükaya for their assistance during the review of the article and subject.

Disclosures of Conflicts of Interest: **J.C.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: consultant for Guerbet, Philips Healthcare, and Eisai; institution receives grants from Guerbet, Philips Healthcare, Boston Scientific, and BTG; author receives travel accommodations from Guerbet and Philips Healthcare. Other relationships: disclosed no relevant relationships. **J.S.D.** Activities related to the present article: institution receives grant from National Institutes of Health, funding from NCI for research in image analysis related to liver cancer diagnosis and treatment. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships.

References

1. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018;69(1):182–236 [Published correction appears in *J Hepatol* 2019;70(4):817.] <https://doi.org/10.1016/j.jhep.2018.03.019>.
2. Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study. *Liver Int* 2015;35(9):2155–2166.
3. Morshid A, Elsayes KM, Khalaf AM, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell* 2019;1(5):e180021.
4. Abajian A, Murali N, Savic LJ, et al. Predicting Treatment Response to Intra-arterial Therapies for Hepatocellular Carcinoma with the Use of Supervised Machine Learning—An Artificial Intelligence Concept. *J Vasc Interv Radiol* 2018;29(6):850–857.e1.
5. Bello GA, Dawes TJW, Duan J, et al. Deep learning cardiac motion analysis for human survival prediction. *Nat Mach Intell* 2019;1:95–104.
6. Zhuang J, Dvornik NC, Li X, Ventola P, Duncan JS. Prediction of Severity and Treatment Outcome for ASD from fMRI. In: Rekik I, Unal G, Adeli E, Park S, eds. *Predictive Intelligence in Medicine*. PRIME 2018. Lecture Notes in Computer Science, vol 11121. Cham, Switzerland: Springer, 2018 https://doi.org/10.1007/978-3-030-00320-3_2.