Perspective

Toward Augmented Radiologists: Changes in Radiology Education in the Era of Machine Learning and Artificial Intelligence

Shahein H. Tajmir, MD, Tarik K. Alkasab, MD, PhD

Radiology practice will be altered by the coming of artificial intelligence, and the process of learning in radiology will be similarly affected. In the short term, radiologists will need to understand the first wave of artificially intelligent tools, how they can help them improve their practice, and be able to effectively supervise their use. Radiology training programs will need to develop curricula to help trainees acquire the knowledge to carry out this new supervisory duty of radiologists. In the longer term, artificially intelligent software assistants could have a transformative effect on the training of residents and fellows, and offer new opportunities to bring learning into the ongoing practice of attending radiologists.

Key Words: Artificial intelligence; education; machine learning.

INTRODUCTION

The machine learning (ML) technique offers new abilities to create artificially intelligent (AI) software tools capable of autonomously finding patterns in large datasets and underlies many large-scale software products such as Google Translate, Alexa, and Facebook (1–3). Researchers have succeeded in applying these techniques to medical applications, with recent successes in detecting diabetic retinopathy (4), identifying malignant melanomas (5), and detecting large vessel occlusion in stroke (6,7). Although there has been much discussion in the lay press about the role of ML-based AI tools in radiology—including proclamations that we should stop training radiologists (8)—both AI/ML and medical imaging experts predict that these new software tools will be central to radiologists' practice across the research, clinical, and education domains (1).

Two large domains of software tools using ML techniques are likely to begin to affect the practice of radiology. The first domain consists of computer vision AI systems, which will likely perform three main tasks within medical imaging: classification, segmentation, and extraction of new biomarkers from raw image data. Radiographic bone age (9,10) and brain hemorrhage detection networks (11–13) are examples of classification problems that have been successfully approached with machine learning tools. ML-trained tools for segmentation tasks try to extract a region of interest automatically, such as the left ventricular cavity (14,15) or fat, muscle, and bone in body composition analysis (16). AI software also allow us to process images in ways that would be infeasible for humans, such as deriving dual-energy x-ray absorptiometry (DEXA) scores from routine clinical computed tomography (CT) examinations (17) or calculating organ-specific radiation dose estimates (18). The second domain is natural language processing (NLP)—the ability of software tools to understand human language. The importance of written and spoken language embedded within the practice of radiology suggests another avenue where radiologists' work will be affected by these new technologies, much as voice recognition (VR) technology has transformed the process of radiology report creation over the past 2 decades.

Our purpose is to suggest paradigms for how radiologists should approach these new tools as they are developed and are deployed across the clinical enterprise, paying special attention to the potential short-term and long-term effects of machine learning on radiology education.

AI TOOLS AS NEW PULSE SEQUENCES

The successful deployment of ML tools will require integration across image acquisition, archival, and interpretation. Beyond the creation of the algorithms that can perform basic image analysis tasks, these tools must be selectively

Acad Radiol 2018; ■:■■–■■

From the Department of Radiology, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114 (S.H.T., T.K.A.); Harvard Medical School, Boston, Massachusetts (S.H.T., T.K.A.); MGH/BWH Center for Clinical Data Sciences, Boston, Massachusetts (T.K.A.). Received February 21, 2018; revised March 7, 2018; accepted March 8, 2018. **Address correspondence to:** T.K.A. e-mail: TALKASAB@mgh.harvard.edu

https://doi.org/10.1016/j.acra.2018.03.007

integrated into the clinical workflow so that radiologists can leverage their power to improve the clinical care of patients. This means that the prospective users—radiologists—will need to be educated on algorithm capabilities and shortcomings. Successful integration will require radiologists to provide input at all levels of the imaging chain and will be similar to the clinical deployment of a new scanner modality. We propose three levels of knowledge that radiologists will need, which are akin to the roles involved in deploying magnetic resonance technology.

The most in-depth knowledge base will be that of the AI tool creators. These physician-scientists will have strong computer science and data science backgrounds to create and to evaluate neural networks and other software tools, and apply those techniques to clinical problems, much as magnetic resonance (MR) researchers design new pulse sequences. The next level of understanding will be the AI tool deployers. These are the radiologists who may not have the technical capability to create new tools, but understand the core concepts well enough to know which systems will interact well with their scanner fleet and patient population, allowing them to create appropriate "protocols" for assisted or autonomous image interpretation. They will require a new skill set for the evaluation of AI tools, both when considering new tools and when evaluating the ongoing performance of deployed tools.

Lastly, the most superficial understanding will be among the AI-using radiologists. These users will not create or design new tools, but will have to know how to appropriately match AI tools to patients and pathologies, similar to how even "physics-averse" radiologists generally know that T1-weighted sequences are used to define anatomy, T2-weighted sequences highlight fluid, and a diffusion-weighted series highlight acute infarction. In addition, front-line radiologists should be able to recognize circumstances under which the deployed AI tools are likely to fail. All residents will need to attain at least baseline competence as AI users, just as they are expected to become competent in operating picture archiving and communication systems (PACS) and voice-dictation systems. They will need to judge whether tool output is plausible and meaningful before relying on it.

AI IN ACADEMIC DEPARTMENTS IN THE SHORT TERM

Most tools available from both new and established vendors offer a vision of "augmented" intelligence. These tools help predictate and preanalyze examinations so that by the time a radiologist opens the study, many tedious detection, measurement, and description tasks have already been performed. For example, one vendor claims to offer a "virtual resident" that will predictate chest CTs, identifying up to five pulmonary nodules and predrafting report text with nodule sizes and image numbers (19). Radiologists are taking on a new responsibility on top of image acquisition and interpretation: active supervision of AI tools. Therefore, trainees will also need to learn how to step into that role.

This new curriculum will need to foster an understanding of how ML-based algorithms work so that trainees are able to judge when tools are applicable to clinical situations and evaluate whether a given tool is working as expected in their practice. Much as residents are taught enough physics to understand image acquisition artifacts, they will need to be taught enough about data science and ML-based AI in particular to recognize "artifacts" such as overfitting or incomplete training. Residents will have to learn to ask the question, "Given this patient and these images (acquired on this machine), is the output of this AI tool usable?"

Training programs will also be faced with the problem of when to introduce new AI tools into trainees' clinical work. For example, our department deployed an automated radiographic bone age assessment tool into the pediatric radiology clinical workflow. The majority of residents that rotate on pediatric imaging are inexperienced first-year residents, raising the question whether residents should be expected to perform manual bone age assessment competently before being allowed to use the tool (which further begs the question of assessing competency). The bone age scenario provides a straightforward measure as to whether the system output is completely erroneous, because our system's output is both the bone age and the representative Greulich and Pyle atlas image. If the output is significantly divergent, the atlas image will not look anything like the patient's hand. By contrast, these same concerns could apply for automated ventricular segmentation in cardiac imaging or visceral adipose tissue quantification, where the outputs are numeric and harder to intuitively verify. It is not so easy to verify if the system accurately segmented the left ventricle without actually hand-segmenting the ventricle, ultimately eliminating the time-savings of the AI tool. Perhaps standardized teaching cases where the AI tools perform well and poorly will be developed to assess a resident's ability to know when to leverage an AI tool's output and when to discard them because of failure.

Many of the first-generation AI tools are being marketed toward the use case of a "virtual resident," an assistant to make an attending radiologist's work easier. However, training programs will need to ensure actual residents and fellows are not cut out of the interpretive loop on both normal and abnormal examinations simply because an AI system has already predictated the case. Residents should be still expected to leverage the tools and predictate cases so they have firsthand experience with the new workflow and also know how and why these systems can fail. On the other hand, residency programs should not use the presence of residents as a reason to avoid adopting the new tools, simply because they have actual residents. To do so would be a failure to educate their residents in an important new aspect to radiology practice, akin to delaying the adoption of MR because CT was "good enough." As the challenges of integrating new AI tools into clinical radiology workflow are addressed, this facet of academic radiology's mission will have to be accounted for, perhaps by including a "training mode" for the integration of new tools.