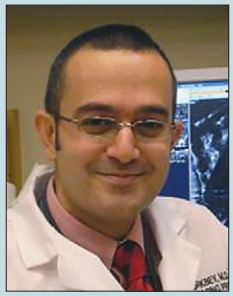
Radiology

Detecting Prostate Cancer with Deep Learning for MRI: A Small Step Forward

Anwar R. Padhani, MRCP, FRCR • Baris Turkbey, MD

Dr Anwar R. Padhani is a radiologist and head of imaging research at the Paul Strickland Scanner Centre, Mount Vernon Cancer Centre, London. He is a professor of cancer imaging at the Institute of Cancer Research, London. He is a member of the executive board and trustee of the International Cancer Imaging Society (ICIS). He is co-chair of the International Prostate Imaging Reporting and Data System (PI-RADS) Steering Committee.

Dr Baris Turkbey is an associate research physician at the Molecular Imaging Program, National Cancer Institute, National Institutes of Health. He is a member of the International Prostate Imaging Reporting and Data System (PI-RADS) Steering Committee. His main research areas include imaging of prostate cancer (multiparametric MRI, PET/CT), prostate biopsy techniques, focal therapy of prostate cancer, and artificial intelligence.



In recent years, high-level evidence of the use of MRI for diagnosis of prostate cancer has emerged for both biopsy-naïve patients and those with negative results from a prior biopsy. This evidence is summarized in the influential Oxford Cochrane review (1). The major benefits of the MRI-directed pathway in biopsy-naïve men are reductions in the number of men needing biopsies, the number of biopsy cores used to make diagnoses, and in decreased detection rates of indolent cancers. In patients with negative results from prior biopsy, increased diagnoses of clinically significant cancer is an additional benefit (2).

The Prostate Imaging Reporting and Data System, or PI-RADS, standard for multiparametric MRI evaluation and reporting (3) has been adopted into guidelines for diagnosis of prostate cancer (4,5). As a result, it is anticipated that there will be rapid adoption of MRI and MRI-guided biopsy for diagnosis of prostate cancer worldwide. To deliver the benefits of the MRI-diagnosis pathway, there is an important need to increase work efficiency while minimizing variations in MRI data acquisition and reader interpretations, and also to decrease the number of data processing steps to identify men who are likely to have clinically significant cancers (2).

The delivery of patient benefits is dependent on reader expertise. Expert readers make more accurate diagnoses with less uncertainty. High level of expertise also enables the adoption of MRI approaches that avoid contrast medium injections (biparametric MRI) (6,7), helping to reduce costs and increase patient throughput. Reader expertise also minimizes variations in clinically significant cancer yields within the PI-RADS suspicion categories, thus improving uniformity and reliability of MRI findings for clinical decision making. High-quality prostate gland and within-gland target lesion delineations are essential to direct MRI-guided biopsies and therapy planning.

The achievement of patient benefits is challenging to deliver because of increasing demands on the radiologist's time and a workforce not fully trained to interpret prostate MRI. There is a steep learning curve for prostate MRI interpretations. Currently, there is wide variation in MRI interpretative performance. Biopsy target detection and delineations in preparation for biopsy is also time consuming. Artificial intelligence, or AI, systems that use deep learning approaches are potentially helpful for automating multiple steps in diagnosis of prostate cancer at MRI (8,9).

A helpful AI system should be able to segment the prostate gland for biopsy and radiation therapy guidance, even when images are not ideal. In the next step, lesions suspicious for prostate cancer should be detected, graded for the likelihood of malignancy, and given probabilities of aggressiveness. Delineation of lesions suspicious for cancer with low false-positive results and high similarity to actual tumor locations will assist in planning biopsy and focal therapies and boost radiation therapy indications. Experience shows that object detection and classifications need plentiful well-annotated data, provided by experts, for training AI systems. These data are difficult to find because digital annotations of prostate MRI examinations are rarely performed in routine clinical practice, and histopathologic correlation with whole-gland tissue and tumor mapping for MRI verification does not happen in the clinical routine when men undergo MRI for suspected cancers.

In this issue of *Radiology,* Schelb et al (10) demonstrate the potential of deep learning approaches to provide diagnostic support to radiologists for image interpretation and target delineation by using diagnostic biparametric MRI. The current work included a curated sample of 250 patients, representative of a mixed outpatient urologic clinic (biopsy-naïve patients, patients with negative findings from prior biopsy, and patients treated with active surveillance). Although this is a relatively small sample size, with training, the AI algorithm had similar performance to that of eight radiologists who had interpreted the multiparametric MRI scans

From the Paul Strickland Scanner Centre, Mount Vernon Cancer Centre, Rickmansworth Road, Northwood, Middlesex HA6 2RN, England (A.R.P.); and Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, Md (B.T.). Received September 4, 2019; revision requested September 9; final revision received September 10; accepted September 16. **Address correspondence to** A.R.P. (e-mail: *anwar.padhani@stricklandscanner.org.uk*).

Conflicts of interest are listed at the end of this article

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

Radiology 2019; 293:618–619 • https://doi.org/10.1148/radiol.2019192012 • Content codes: GU MR •

by using the PI-RADS system (3) and to one investigator who performed prostate gland and intraprostate lesion delineations. Histologic validation used a transperineal mapping biopsy scheme known for its high accuracy.

Of note, the success of the AI came at the cost of a high false-positive rate of around 50%. The sensitivity of the AI system rapidly declined when attempting to lower the false-positive rate. That is, to maintain high sensitivity for the detection of clinically significant cancer, the AI overcalls multiple lesions that do not represent significant cancers (false-positive findings). A high false-positive rate of AI systems has been noted by other investigators (8,9). Clinical tolerance to false-positive results differs between patient groups (2). In biopsy-naïve men, there is a need to minimize overdiagnoses and to detect significant cancers, therefore reducing the acceptance of false-positive results. However, management priority in men with persistent suspicion of clinically significant cancers after negative findings at previous biopsy is to not miss potentially life-threatening cancers, therefore increasing tolerance for false-positive results.

In the short term, the disadvantage of high false-positive rates can be addressed by using AI systems as triage tools that detect and present lesions suspicious for cancer, along with their delineations, to radiologists. In a second step, the radiologists can either accept or reject proposed lesions in line with clinical risk factors and patient priorities, and improve delineations of targets suspicious for cancer before proceeding to reporting and communication tasks. It is interesting to note that when radiologists and machine learning agree on likely presence of clinically significant lesions, the positive predictive value increases without affecting the negative predictive value.

A positive result of Schelb et al (10) was the high similarity of the AI algorithm to segment the entire gland comparable to manual segmentations. This opens the possibility for segmentation of the prostate gland to be used immediately for fusion biopsy and radiation therapy planning.

The performance of the deep learning system described by Schelb et al (10) is expected to improve when larger data sets are used for training. It is difficult to benchmark the collective performance of the radiologists' readings given the large number of readers, the limited number of scans evaluated per reader, and the mixed population of patients. The similarity scores for detected lesions are relatively low, and we do not get a strong sense about lesion similarity scores and measurements of tumor size or aggressiveness. However, we must acknowledge that this is a hard task and is subject to the variability of lesion outlining by one of the investigators, remembering that there is no recognized operating procedure on how lesion outlining should be performed. Bringing objectivity to the outlining process for AI training would require histopathologic correlations with both prostatectomy specimens and transperineal template mapping biopsies using high spatial sampling densities (because the entire population presenting for diagnosis must be represented in the training data set). Finally, it is important to note that this is a limited single-center data set, and we have no information on how the AI system would operate with other MRI protocols and scanners. Furthermore, we cannot directly extrapolate these results to men with much higher or much lower risks of having clinically significant cancers.

In the next phases of optimization of AI systems, larger, well-curated, and diverse (potentially from multiple vendors, multiple centers) training data sets with spatially well-correlated histopathologic validation must be used. Although such developments can be time and resource consuming, AI systems for diagnosis of prostate cancer at MRI will certainly perform better, although we do not yet know what the performance will level out at compared with trained human readings. Ideally, AI systems should be tested in prospective randomized studies to assess test performance, from which robust measurements of clinical impacts can be derived in different use case scenarios.

AI systems that can detect normal cases with high confidence can enable the identification of patients who will likely not need biopsies, thus optimizing reading workflow. A second reader role for indeterminate MRI cases that includes integrations with clinical history, biochemistry, and genomic profiles can also be envisioned. Clearly, AI systems will need to be tuned to population disease prevalence, and to have performance characteristics that enable the delivery of diagnostic benefits according to individual patient clinical priorities. These steps are essential to developing community-wide expertise in diagnosis of prostate cancer at MRI. Robust standard operating procedures will increase confidence of patients and payers, enabling the wider adoption of the MRI-directed approach. We recommend that radiologists engage in the clinical development of AI systems for diagnosis of prostate cancer to meet the increasing demands for MRI-directed diagnosis of prostate cancer.

Disclosures of Conflicts of Interest: A.R.P. disclosed no relevant relationships. **B.T.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has cooperative research and development agreements with Philips and Nvidia; receives royalties from Invivo. Other relationships: has patent for related intellectual property in field of prostate computer-aided diagnosis (National Institutes of Health–owned).

References

1. Drost FH, Osses D, Nieboer D, et al. Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. Eur Urol 2019 Jul 17 [Epub ahead of print].
2. Padhani AR, Barentsz J, Villeirs G, et al. PI-RADS Steering Committee: The PI-RADS Multiparametric MRI and MRI-directed Biopsy Pathway. Radiology 2019;292(2):464–474.
3. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. Eur Urol 2019;76(3):340–351.
4. Mottet N, van den Bergh RC, Briers E, et al. 2019 EAU - EANM - ESTRO - ESUR - SIOG Guidelines on Prostate Cancer. EAU Guidel. Edn. Present. EAU Annu. Congr. Barcelona 2019. Arnhem, the Netherlands: EAU Guidelines Office, 2019. https://uroweb.org/guideline/prostate-cancer/#3.
5. NCCN Clinical Practice Guidelines in Oncology V1.2019. Prostate Cancer Early Detection Recommendations. Natl Compr Cancer Network, Inc. 2019. https://www.nccn.org/professionals/physician_gls/pdf/prostate_detection.pdf. Accessed DATE.
6. van der Leest M, Israël B, Cornel EB, et al. High Diagnostic Performance of Short Magnetic Resonance Imaging Protocols for Prostate Cancer Detection in Biopsy-naïve Men: The Next Step in Magnetic Resonance Imaging Accessibility. Eur Urol 2019 Jun 2 [Epub ahead of print].
7. Gatti M, Faletti R, Calleris G, et al. Prostate cancer detection with biparametric magnetic resonance imaging (bpMRI) by readers with different experience: performance and comparison with multiparametric (mpMRI). Abdom Radiol (NY) 2019;44(5):1883–1893.
8. Gaur S, Lay N, Harmon SA, et al. Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? a multi-center, multi-reader investigation. Oncotarget 2018;9(73):33804–33817.
9. Giannini V, Mazzetti S, Armando E, et al. Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. Eur Radiol 2017;27(10):4200–4208.
10. Schelb P, Kohl S, Radtke JP, et al. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. Radiology 2019;293:607–617.