Artificial intelligence for prostate cancer histopathology diagnostics

Victor Sandoval[1], Zachary Chuang[1], Nicholas Power[2,3], Joseph L.K. Chin[2,3]

[1]Schulich School of Medicine & Dentistry, Western University, London, ON, Canada; [2]Department of Surgery, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada; [3]Department of Oncology, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

Cite as: Sandoval V, Chuang Z, Power N, et al. Artificial intelligence for prostate cancer histopathology diagnostics. *Can Urol Assoc J* 2022;16(12):439-41. http://dx.doi.org/10.5489/cuaj.7918

Published online July 21, 2022

Recently, there has been significant interest in the application of artificial intelligence (AI) technology to cancer diagnostics. In uro-oncology, this is evident by the significant growth in publications focusing on AI and prostate cancer (PCa) histopathology.[1] Recent advancements in digital and computer vision technologies have the potential to revolutionize the diagnosis and grading of PCa. In conjunction with well-designed AI models, core prostate biopsy imaging and whole slide imaging (WSI) techniques could lead to quicker, more reliable and exact diagnoses.[1,2] These advancements would increase automation and provide diagnostic standardization, with the added benefit of reducing workloads on overburdened pathology departments.[1,2] Despite this immense potential, many questions have been raised about the feasibility and clinical applicability of these technologies in urologic oncology.

The application of AI in PCa diagnostics has largely focused on machine learning (ML) — a branch of AI based on the development and training of algorithms with the ability to learn from historical data inputs, without explicitly programming a set of matching outputs to inputs.[3] Deep learning (DL) is a subcategory of ML based on artificial neural networks that mimic the function of human neurons. Like the feature detection system of the sensory nervous system, DL uses a multilayer approach to progressively extract features from raw input, making it particularly facile with image processing.[3] Digitization has allowed groups to create foundational reference sets for AI models with thousands of data points linking images with their respective expert-assigned Gleason scores.[4-6] From this backdrop, AI machines could be used to develop grading outputs that reduce pathologist inter-variability and increase diagnostic accuracy.

The ability to distinguish between benign and malignant disease is fundamental in pathology, thus representing a requisite capability for any serviceable AI model.[6,7] Campanella et al demonstrated that AI models can identify malignancy and exclude benign tissue samples for various cancers (including PCa) with extraordinarily high sensitivity.[5] Further, a study conducted by Han et al compared seven different AI models and found that all were capable of making the determination between cancerous and non-cancerous tissues, with error rates of only 6–14%.[8] These findings suggest that several ML and DL models could be used to automatically screen pathology samples to identify benign slides. These samples could subsequently be excluded, while those determined to be suspicious or malignant could be sent for formal or secondary review by a pathologist. Applying the technology in this fashion would triage incoming samples, leading to more effective resource utilization.

Beyond the determination of malignancy, the ideal AI model needs to be able to make distinctions between high- and low-grade disease.[6,7] This has already been accomplished by some groups, including Silva-Rodriquez et al. They reviewed 6682 digitalized prostate biopsy cores using an AI model that automatically supported the pathologist's analysis of WSI using cribriform pattern detection. This model was designed to identify cribriform architecture in Gleason 4, which is associated with adverse prognostic features that imply higher-risk disease. Their model demonstrated excellent pattern discrimination, with overall performance similar to that of general pathologists.[9] In another study, a similar AI model was employed to detect these cribriform patterns in prostate needle biopsies. This model achieved a sensitivity of 0.9, with limited false-positives.[10] Despite these successes, it is important to note that different AI models vary in their grading accuracy. For example, it has been shown that models employing tissue component maps (TCMs) outperform those that use raw inputs, particularly when analyzing the most aggressive PCa tissue types (i.e., Gleason 5).[11] Models, such as the aforementioned, demonstrate utility *beyond* the triage setting and showcase their potential to assist pathologists with grading through the detection and identification of suspicious architectures.

Subspecialized genitourinary (GU) pathologists represent the gold standard in histopathological analysis of PCa; however, not every center has the resources to employ these

niche and expertly trained professionals. General pathologists typically serve as the backbone in many departments due to their broad knowledge base and clinical flexibility. Some groups have suggested that AI technology could serve as excellent adjunct tools to improve the diagnostic capability of general pathologists.[9] For example, Nagpal et al compared the rate of diagnostic agreement between a DL system with both general and GU pathologists. The overall rate of agreement, using digital images of prostate tissues samples, was approximately 95% in the diagnosis of malignancy for all groups; however, the DL system outperformed general pathologists in the Gleason scoring of malignant specimens. The AI model obtained a 71.7% rate of agreement with GU pathologists compared to a 58% rate of agreement for general pathologists.[12] Bulten et al developed a similar DL model, which also showed high agreement with their reference standard — a data set that had been developed by expert consensus. In fact, the DL system performed comparably to pathologists with more than 15 years of experience and managed to outperform those with less than 15 years' experience.[7] Taken together, these findings clearly demonstrate the ability of these models to address gaps in both service and experience, with some utility in supporting clinical decisions.

Despite the increasing demand for pathology services, there has been a decline in the number of practicing pathologists in recent years.[13] Given that this phenomenon is complex and multifactorial,[14] centers may wish to pursue streamlined options that mitigate these human resource deficits. Various studies have shown that AI models can performed reasonably well, despite the relatively early stage of development;[5-12] however, when it comes to the histopathological diagnosis of disease, it is important to recognize that these models are only capable of lending *some* of the expertise of GU pathologists. Hence, current AI models would not be capable of outperforming the diagnostic abilities or substitute for the clinical acumen of these highly trained subspecialists.

The ability for AI models to learn from previous data and create new outputs is unquestionably fascinating. Studies suggest that the implementation of this technology may bring novelty and improved accuracy to the diagnosis of PCa by providing standardization in Gleason scoring; however, the development of this technology and the performance of rigorous testing takes incredible amounts of resources, including knowledge, time, and money. While this technology has the potential to provide practical, clinical, and financial value to the resource-rich centers that employ them, it may simply be unattainable for resource-challenged institutions and jurisdictions.

Additional concerns have been raised about the generalizability of AI studies, particularly when considering the applicability of AI technology to patients at non-academic sites (i.e., community hospitals), multinational cohorts, or those of minority subgroups.[15] Since most studied AI models are developed using data from a single academic center, they risk inherent bias from their respective patient population.[16] The recent Prostate cANcer graDe Assessment (PANDA) challenge sought to address these limitations as the largest histopathology competition to date; the goal was to catalyze the development of reproducible AI algorithms using over 10 000 multicenter, multinational, digitized prostate biopsies.[16] They validated algorithms that achieved an incredible 0.862 κ and 0.868 κ with expert uropathologists on United States and European external validation sets, respectively.[16] While this finding certainly shows promise, AI output for foreign cases, such as histological variants or samples with chronic inflammation, is still unknown. Fortunately, it may be possible to mitigate the risk of critical predictive errors using an algorithmic audit to identify potential blind spots within an AI model prior to clinical deployment.[17] Beyond multinational competitions and AI auditing, continued efforts to establish generalizability, such as adequately powered studies (i.e., adequate sample sizes) and the adoption of standardized reporting in AI research (using guidelines such as the Radiomics Quality Score[18] or STREAM-URO framework[19]), should be encouraged. Future research could also focus on clinical application, such as the prediction of long-term outcomes, in order to demonstrate utility beyond pure pathological science.

While AI studies in PCa have been promising, they have not been able to demonstrate the superiority of AI models compared to the diagnostic prowess and clinical performance of GU pathologists. Despite the apparent utility of this technology in bridging various gaps associated with general pathology assessment, it is evident that AI models in their current form are not "ready for prime time" nor to supplant our GU pathologists. For AI technology to be accepted, employed, and trusted in the field, more research needs to be done to ensure this technology is widely cost-effective, scalable, reproducible, generalizable, and provides meaningful outputs at a level that exceeds the standard of care. Evidently, this is quite a tall order.

Competing interests: The authors do not report any competing personal or financial interests related to this work.

This paper has been peer-reviewed.

References

1. Elkhader J, Elemento O. Artificial intelligence in oncology: From bench to clinic. *Sem Cancer Biol* 2021;84:113-28. https://doi.org/10.1016/j.semcancer.2021.04.013
2. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J Int Med* 2020;88:62-81. https://doi.org/10.1111/joim.13030

3. Goldenberg SL, Nir G, Salcudean SE. A new era: Artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019;16:391-403. https://doi.org/10.1038/s41585-019-0193-3
4. Chin J, Bauman G, Power N, et al. The singularity is near(ish): Emerging applications of artificial intelligence in prostate cancer management. *Eur Urol* 2020;77:293-5. https://doi.org/10.1016/j.eururo.2019.12.006
5. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019;25:1301-9. https://doi.org/10.1038/s41591-019-0508-1
6. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *Lancet Oncol* 2020;21:222-32. https://doi.org/10.1016/S1470-2045(19)30738-7
7. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *Lancet Oncol* 2020;21:233-41. https://doi.org/10.1016/S1470-2045(19)30739-9
8. Han W, Johnson C, Gaed M, et al. Histological tissue components provide major cues for machine learning-based prostate cancer detection and grading on prostatectomy specimens. *Sci Rep* 2020;10:9911. https://doi.org/10.1038/s41598-020-66849-2
9. Silva-Rodríguez J, Colomer A, Sales MA, et al. Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. *Comp Meth Prog Biomed* 2020;195:105637. https://doi.org/10.1016/j.cmpb.2020.105637
10. Ambrosini P, Hollemans E, Kweldam CF, et al. Automated detection of cribriform growth patterns in prostate histology images. *Sci Rep* 2020;10:14904. https://doi.org/10.1038/s41598-020-71942-7
11. Han W, Johnson C, Warner A, et al. Automatic cancer detection on digital histopathology images of mid-gland radical prostatectomy specimens. *J Med Imag (Bellingham)* 2020;7:047501. https://doi.org/10.1117/1.JMI.7.4.047501
12. Nagpal K, Foote D, Tan F, et al. Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020;6:1372-80. https://doi.org/10.1001/jamaoncol.2020.2485
13. Metter DM CTLS et al. Trends in the U.S. and Canadian pathologist workforces from 2007–2017. *JAMA Netw Open* 2019;2:e194337. https://doi.org/10.1001/jamanetworkopen.2019.4337
14. Robboy SJ, Weintraub S, Horvath AE, et al. Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch Pathol Lab Med* 2013;137:1723-32. https://doi.org/10.5858/arpa.2013-0200-OA
15. Gichoya JW, McCoy LG, Celi LA, et al. Equity in essence: A call for operationalizing fairness in machine learning for healthcare. *BMJ Health Care Inform* 2021;28:100289. https://doi.org/10.1136/bmjhci-2020-100289
16. Bulten W, Kartasalo K, Chen PH, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The PANDA challenge. *Nat Med* 2022;28:154-63. https://doi.org/10.1016/S0302-2838(22)00693-5
17. Liu X, Glocker B, McCradden MM, et al. The medical algorithmic audit. *Lancet Dig Health* 2022;4:e384-97. https://doi.org/10.1016/S2589-7500(22)00003-6
18. Lambin P, Leijenaar RT, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-62. https://doi.org/10.1038/nrclinonc.2017.141
19. Kwong JC, McLoughlin LC, Haider M, et al. Standardized reporting of machine learning applications in urology: The STREAM-URO framework. *Eur Urol Focus* 2021;7:672-82. https://doi.org/10.1016/j.euf.2021.07.004

Correspondence: Dr. Nicholas Power, Departments of Surgery and Oncology, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada; nicholas.power@lhsc.on.ca