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

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

***

Recently, there has been significant interest in the application of 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 multi-layer approach to progressively extract features from raw input, making it particularly facile with image processing. (3) Digitization has allowed groups to create foundational references 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) Incredibly, this feat 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 whole block imaging (WBI) using cribriform pattern detection. This model was designed to identify cribriform architecture in Gleason pattern 4 PCa, 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. (10) 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. (11) 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). (12) 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.

Sub-specialized Genito-urinary Pathologists ("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. (10) 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

CUAJ – **Community Perspective**
                                                    Sandoval et al
                                        AI for prostate cancer histopathology diagnostics

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. (9) Bulten et al. developed a similar DL model, which also showed high agreement with their reference standard; a data set which 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 sub-specialists.

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 multi-center, multi-national 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 pathologic 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.

References

1. Elkhader J, Elemento O. Artificial intelligence in oncology: From bench to clinic. Seminars in Cancer Biology. Academic Press; 2021.
2. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. Vol. 288, Journal of Internal Medicine. Blackwell Publishing Ltd; 2020. p. 62–81.
3. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Vol. 16, Nature Reviews Urology. Nature Publishing Group; 2019. p. 391–403
4. Chin J, Bauman G, Power N, Ward A. The Singularity is Near(ish): Emerging Applications of Artificial Intelligence in Prostate Cancer Management. Vol. 77, European Urology. Elsevier B.V.; 2020. p. 293–5.
5. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019 Aug 1;25(8):1301–9.
6. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. The Lancet Oncology. 2020 Feb 1;21(2):222–32.
7. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. The Lancet Oncology. 2020 Feb 1;21(2):233–41.
8. Han W, Johnson C, Gaed M, Gómez JA, Moussa M, Chin JL, Pautler S, Bauman GS, Ward AD. Histologic tissue components provide major cues for machine learning-based prostate cancer detection and grading on prostatectomy specimens. Sci Rep, 2020 Jun 18; 10 (1): 9911.
9. Nagpal K, Foote D, Tan F, Liu Y, Chen PHC, Steiner DF, et al. Development and Validation of a Deep Learning Algorithm for Gleason Grading of Prostate Cancer from Biopsy Specimens. JAMA Oncology. 2020 Sep 1;6(9):1372–80.
10. Silva-Rodríguez J, Colomer A, Sales MA, Molina R, Naranjo V. Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. Computer Methods and Programs in Biomedicine. 2020 Oct 1;195
11. Ambrosini P, Hollemans E, Kweldam CF, Leenders GJLH van, Stallinga S, Vos F. Automated detection of cribriform growth patterns in prostate histology images. Scientific Reports. 2020 Dec 1;10(1).
12. Han W, Johnson C, Warner A, Gaed M, Gomez JA, Moussa M, Chin JL, Pautler S, Bauman G, Ward AD. Automatic cancer detection on digital histopathology images of mid-gland radical prostatectomy specimens. J Med Imaging (Bellingham). 2020 Jul;7(4):047501. doi:
13. Metter DM CTLS et al. Trends in the US and Canadian Pathologist Workforces from 2007 to 2017 . JAMA Netw Open . 2019;2.
14. Robboy SJ, Weintraub S, Horvath AE, Jensen BW, Alexander CB, Fody EP, et al. Pathologist workforce in the United States: I. Development of a predictive model to

examine factors influencing supply. Archives of Pathology and Laboratory Medicine. 2013 Dec;137(12):1723–32.

15. Gichoya JW, McCoy LG, Celi LA, Ghassemi M. Equity in essence: a call for operationalising fairness in machine learning for healthcare. BMJ health & care informatics. 2021;28(1)

16. Bulten W, Kartasalo K, Chen PH, Ström P, Pinckaers H, Nagpal K, Cai Y, Steiner DF, van Boven H, Vink R, Hulsbergen-van de Kaa C. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nature medicine. 2022 Jan 13:1-0

17. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. The Lancet Digital Health. 2022 Apr 5

18. Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, Sanduleanu S, Larue RT, Even AJ, Jochems A, van Wijk Y. Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology. 2017 Dec;14(12):749-62

19. Kwong JC, McLoughlin LC, Haider M, Goldenberg MG, Erdman L, Rickard M, Lorenzo AJ, Hung AJ, Farcas M, Goldenberg L, Nguan C. Standardized Reporting of Machine Learning Applications in Urology: The STREAM-URO Framework. European Urology Focus. 2021 Jul 1;7(4):672-82