

Commentary: Automated Diagnosis and Gleason Grading of Prostate Cancer – Are Artificial Intelligence Systems Ready for Prime Time?

Received: 26 September 2019

Accepted: 25 October 2019

Published: 23 December 2019

The field of pathology is transforming with the emergence of new tools and technologies such as digital and computational pathology and artificial intelligence (AI).^[1-3] This is also an exciting time in advanced diagnostics with the more widespread use of digital imaging in pathology, in particular, the validation and regulatory approvals and increasing use of whole-slide imaging (WSI) technology.^[4] WSI allows the scanning of the entire glass slides, with an outputting of an image file that is a digitized reproduction of the glass slide with images that boast diagnostic quality standards.^[4] In addition, in the past 5 years, we have witnessed an increasing use of machine learning (ML), deep learning (DL), and AI tools in clinical and translational pathology.^[1,5]

The recently published article about the Gleason grading of prostate cancer by Nagpal *et al.* has received significant attention. The authors of this study aimed to build a DL-based scoring tool for the prostate whole-slide sections used in clinical workflows that could help (1) facilitate the reduction in Gleason grading variability and maybe make the Gleason score more objective, (2) provide a way to better risk stratify prostate cancer patients, and (3) improve the management of prostate cancer patients.^[6] The million dollar question is mainstream pathology ready to use these tools in everyday clinical sign-out and is the adoption of an automated Gleason grading system ready for prime time?

GLEASON GRADING OF PROSTATE CARCINOMA

Gleason grading of prostate carcinoma was designed in the early 1960s by the late Dr. Gleason, and it continues to play a pivotal role in the prognostication of prostate cancer patients. Gleason grading remains an important approach to the histological grading of carcinoma of the prostate.^[7] In recent years, there have been a number of new pathological and genetic discoveries as well as modifications in prostate cancer screening and detection methods that have resulted in a need for revision of the original grading system.^[7] Since the 2004 WHO classification, there have been modifications to the Gleason grading system, and these were incorporated into the 2016 WHO section on grading of prostate cancer. In addition, for Gleason score 7 adenocarcinomas, reporting percentage of adenocarcinoma that is pattern Grade 4 is recommended. The new WHO classification system also recommends utilizing the recently developed prostate cancer grade grouping with five grade groups.^[7]

DIAGNOSTIC VARIABILITY IN GLEASON GRADING OF PROSTATE CANCER

Diagnostic errors do occur in pathology, and there are a significant number of cases where pathologists disagree on a diagnosis. Although errors in pathology can occur for several reasons, interobserver variability is one often cited.^[8] In most cases, where a specialist is one of the reviewers, the specialist's interpretation prevails. A more recent review by Peck *et al.* on diagnostic errors in anatomical pathology and the role and value of second opinions in error prevention revealed that based on the results of literature review and assessments from 2015 to 2017, the rate of inaccurate diagnoses ranged from 3% to 9%. The highest mean percentage of inaccurate diagnoses was noted in the subspecialties of gynecology, dermatopathology, and gastrointestinal specimens.^[9]

Discordance and diagnostic variability in prostate cancer diagnosis and grading has also been reported, particularly in Gleason grade diagnosis, when the diagnosis is provided by general pathologists as compared to subspecialist urologic pathologists.^[10,11]

For a pathology practice, the ideal approach to address this variability, reduce misdiagnoses, and improve outcomes is enterprise-wide deployment of digital pathology to enable changes in workflow, whereby the best-equipped/experienced pathologists are reading slides appropriately matched to them. This is much more than simply redirecting high-acuity cases to specialists. Instead, it involves retooling general pathologists into subspecialists by directing increased volumes of specific case types to them in combination with oversight and training by specialists. Further, as the advanced tool sets within digital pathology and AI are developed, such as an AI-based interpretation of H and E-stained slides and ultimately automated Gleason grading/scoring, diagnostic quality will further improve.^[1,6]

AUTOMATED GRADING OF PROSTATE CANCER

Because there is a significant inter-observer variability in histopathological grading of prostate cancer and classifying them into grade groups,^[12] there have been a number of studies recently exploring the use of AI methods to assist with the grading of prostate cancer. Using computer-assisted methods, it is now possible to objectively grade prostate cancer in histopathological slides to augment the diagnosis of the

pathologist and to improve accuracy and reproducibility as demonstrated in a number of DL-based studies.^[1,6,13]

Nagpal *et al.* have developed a DL model to improve Gleason scoring of prostate cancer from prostatectomies.^[6] A large data set of 112 million pathologist-annotated image patches from 1226 slides was used. The authors deployed a novel DL technique that has achieved an accuracy rate of 0.70, as compared to 0.65 accuracy achieved by 29 general pathologists. In addition, when the results were compared to ground truth provided by expert urologic pathologists, the mean “accuracy” among the 29 general pathologists was only 0.61 on the validation set. Using DL algorithms, there was a significantly higher diagnostic accuracy of 0.70 ($P = 0.002$). The authors concluded that these data may be a better patient risk stratification model to stratify patients for appropriate therapy decisions.^[6]

Another recent study by Lucas *et al.* aimed to design and test an automatic Gleason pattern grading system that could help in automatic grade group determination on prostate needle biopsies.^[13] A convolutional neural network was tested on 96 prostate biopsies from 38 patients. All the biopsies were annotated at a pixel level. Relevant patches were extracted from these annotated images. The results demonstrated that the algorithm was able to discriminate between benign prostate and malignant prostate cancer (Gleason pattern ≥ 3) with an overall accuracy of 92%, with a sensitivity and specificity of 90% and 93%, respectively. This was very striking.^[13] When further analysis was done, the results showed that the algorithm when discriminating Gleason patterns ≥ 4 and Gleason patterns ≤ 3 was able to achieve an accuracy of 90%, with a sensitivity and specificity of 77% and 94%, respectively. When compared to the grading done by an expert urologic pathologist, the algorithm performance was 65% ($\kappa = 0.70$). Due to its clinical significance, Gleason pattern quantitation was compared among the groups: Deep learning system (DLS) had 4%–6% lower mean absolute error than the average pathologist and the DLS predicted the same pattern as the pathologist 97% of the time. Finally, the DLS was able to predict a more gradual transition from well to poorly differentiated by utilizing “fine-grained Gleason patterns,” such as 3.3 or 3.7 instead of the traditional methods of Gleason scoring.^[13]

Nir *et al.* have recently developed algorithms which are supervised on digitized prostate histology slides.^[14] The authors created a unique pipeline for the extraction of several features that included the glandular, cellular, and image-based features. Among the features that were extracted, it was found that intra- and inter-nuclei characteristics were the most important ones for classification. The classifiers were trained on 333 tissue microarray cores which were sampled from 231 radical prostatectomy cases.^[14] The slides from these cases were annotated by six pathologists for different Gleason grades. In this study, when the algorithm was tested on additional 230 digital slides from 56 patients, the overall grading agreement of the classifier with the pathologists was found to be a kappa

of 0.51. When compared to the overall agreements between each individual pathologist and the others, the agreements ranged from 0.45 to 0.62. Overall, this study demonstrated that the supervised algorithm’s performance was in the range that was seen in agreement levels between pathologists. The study overall demonstrates that best results (92% accuracy) are seen in prostate cancer detection and a reduced 79% accuracy in classification of low- and high-grade prostate cancers. Overall, the agreement with expert pathologists was fairly low.^[14]

In a subsequent study, Nir *et al.* have reviewed the appropriate approaches to evaluate AI tools for the automatic grading of prostate cancer from histology images. The group designed a quality improvement study to study the performance of a DL classifier that was evaluated to distinguish benign prostatic glands from Gleason patterns 3, 4, or 5. The initial data were from 231 patients and were used for training the algorithm. The ground truth diagnosis and Gleason grade were obtained from either one expert or multiple experts. The study concluded that as newer AI tools are developed for the grading of prostate cancer and other cancer types, it is critical to utilize annotation data by multiple experts for training and validation of these pattern-recognition algorithms.^[15]

Prostate cancer detection and grading using AI methods will continue to be further explored with the end goal of creating a simple and automated workflow for the practicing pathologists. As discussed, there continues to be significant interest by several groups and several recent computational pathology and AI studies have specifically addressed Gleason grading of prostate cancer. Advances in digital imaging and ability to rapidly digitize glass slides have now paved the way to start using these images in DL/ML algorithms, leading to potentially novel and innovative diagnostic tools.

The use of AI of digital pathology images can extend the value of digital pathology far beyond what is possible today and quantified above. Some barriers such as cost of technology and the ease of integration with information systems need to be overcome for more widespread digitization of pathology. Mainstream pathology practices are now starting to realistically look at these solutions, and pathologists are making more of an effort to learn about these tools and decision support systems. At the crux of this adoption, lies in the question if these systems are good enough and are they ready for prime time? Once this integration journey further matures, that is when the true value of digital pathology and the tools that are proposed by these AI experts will be fully realized. In the meantime, studies such as those described by Nagpal *et al.* will continue to pave the way of discoveries which are much needed to demonstrate the efficiency and concordance of these AI algorithms as compared to the human observers. Eventually, these tools were described by Nagpal *et al.*, and other such studies in the future will help become an important decision support tool for the general and specialist pathologist, leading to true “augmented” and intelligent pathology systems.

Anil V. Parwani¹¹Department of Pathology, The Ohio State University Wexner Medical Centre, Columbus, OH, USA**Address for correspondence:** Dr. Anil V. Parwani, Department of Pathology, Wexner Medical Center, E409 Doan Hall, 410 West 10th Ave, Columbus, OH 43210, USA. E-mail: anil.parwani@osumc.edu

REFERENCES

- Niazi MK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019;20:e253-61.
- Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, *et al.* Computational pathology definitions, best practices, and recommendations for regulatory guidance: A white paper from the digital pathology association. *J Pathol* 2019;249:286-94.
- Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, *et al.* Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. *J Pathol Inform* 2019;10:9.
- Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, *et al.* A practical guide to whole slide imaging: A White paper from the digital pathology association. *Arch Pathol Lab Med* 2019;143:222-34.
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, *et al.* Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137:1710-22.
- Nagpal K, Foote D, Liu Y, Chen PC, Wulczyn E, Tan F, *et al.* Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med* 2019;2:48.
- Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: Current state and prognostic implications. *Diagn Pathol* 2016;11:25.
- Raab SS, Grzybicki DM, Janosky JE, Zarbo RJ, Meier FA, Jensen C, *et al.* Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer* 2005;104:2205-13.
- Peck M, Moffat D, Latham B, Badrick T. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. *J Clin Pathol* 2018;71:995-1000.
- Steinberg DM, Sauvageot J, Piantadosi S, Epstein JI. Correlation of prostate needle biopsy and radical prostatectomy Gleason grade in academic and community settings. *Am J Surg Pathol* 1997;21:566-76.
- Truesdale MD, Cheetham PJ, Turk AT, Sartori S, Hruby GW, Dinneen EP, *et al.* Gleason score concordance on biopsy-confirmed prostate cancer: Is pathological re-evaluation necessary prior to radical prostatectomy? *BJU Int* 2011;107:749-54.
- Goodman M, Ward KC, Osunkoya AO, Datta MW, Luthringer D, Young AN, *et al.* Frequency and determinants of disagreement and error in gleason scores: A population-based study of prostate cancer. *Prostate* 2012;72:1389-98.
- Lucas M, Jansen I, Savci-Heijink CD, Meijer SL, de Boer OJ, van Leeuwen TG, *et al.* Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies. *Virchows Arch* 2019;475:77-83.
- Nir G, Hor S, Karimi D, Fazli L, Skinnider BF, Tavassoli P, *et al.* Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Med Image Anal* 2018;50:167-80.
- Nir G, Karimi D, Goldenberg SL, Fazli L, Skinnider BF, Tavassoli P, *et al.* Comparison of artificial intelligence techniques to evaluate performance of a classifier for automatic grading of prostate cancer from digitized histopathologic images. *JAMA Netw Open* 2019;2:e190442.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

Access this article online

Quick Response Code:**Website:**www.jpathinformatics.org**DOI:**

10.4103/jpi.jpi_56_19

How to cite this article: Parwani AV. Commentary: Automated diagnosis and gleason grading of prostate cancer – are artificial intelligence systems ready for prime time? *J Pathol Inform* 2019;10:41.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2019/10/1/41/273790>