The Reading Paradigm: How the Sequence and Presentation of AI Results to Pathologists Influences Endpoints and Outcomes.

Margaret R. Horton, PhD, Andrea Parke, PhD, Emre Gültürk, MS, Juan A. Retamero, MD, Jill Sue, MS, Brandon Rothrock, PhD, David S. Klimstra, MD

Paige.AI Inc., 11 Times Square, 37 Floor, New York, USA

Introduction

With the increasing availability and market clearances of artificial intelligence (AI)—based algorithms for pathology, there is strong focus on clinical performance and safe use. Consequently, clinical performance studies are evolving from standalone performance studies (human vs. AI) to studies that look at combined human+AI performance.

The sequence and circumstances in which the outputs of the algorithm are presented to pathologists may be referred to as the *reading paradigm.*

Combined human+AI performance studies may be classified according to three categories—second read, concurrent read and pre-screening (Figure 1, using example of prostate cancer detection algorithm). The use of reading paradigm has clear implications on the time spent reviewing cases in clinical practice.

Three Different Reading Paradigms

Example: Prostate Cancer Detection Algorithm

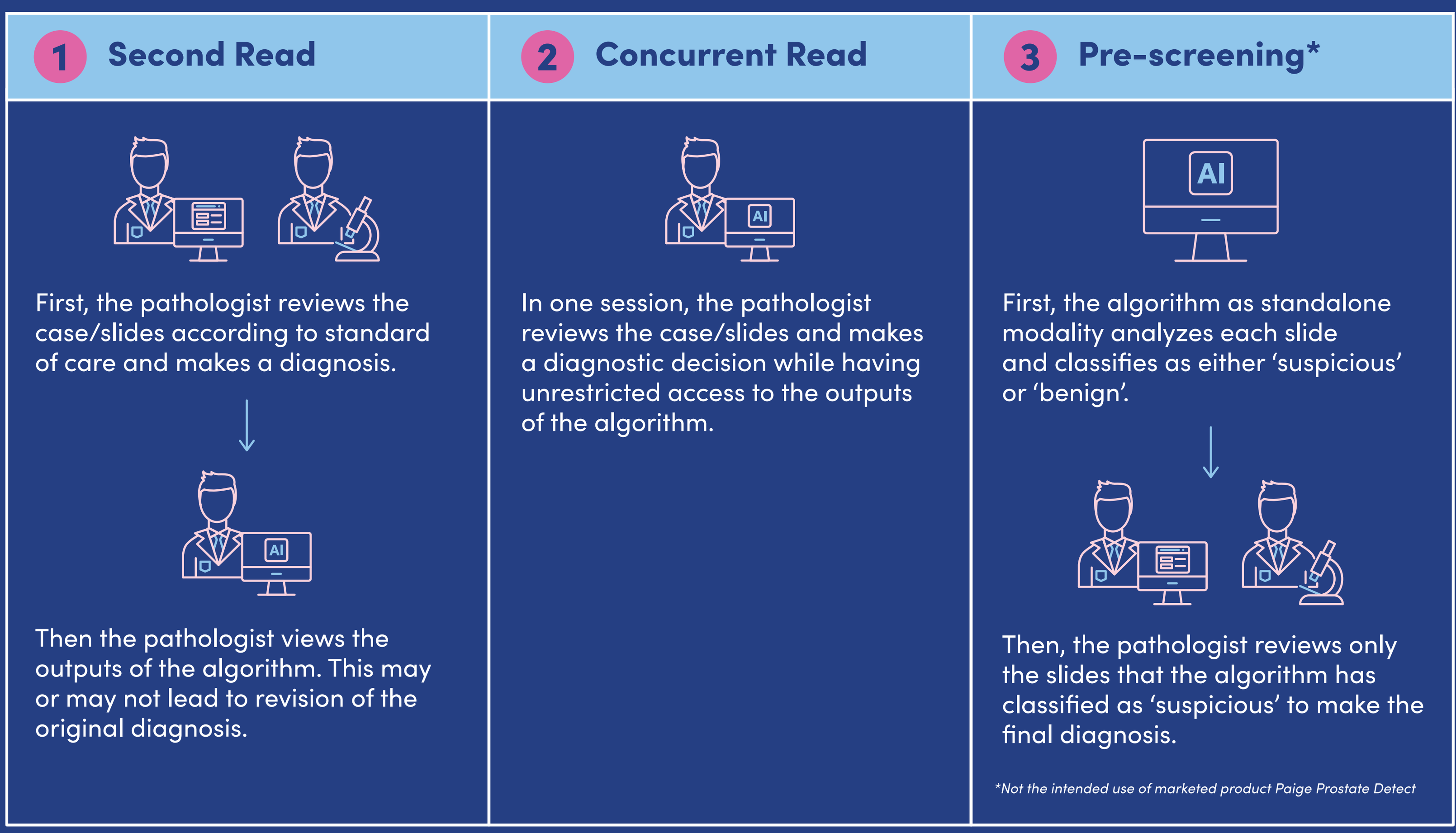

Figure 1. Illustration of three different reading paradigms employed in human+AI performance studies, using the example of AI-based prostate cancer detection. For simplicity, use of ancillary testing (such as immunohistochemical stains, additional levels prepared, second opinions) beyond primary diagnosis of available slides and stains is not represented in the workflow.

Materials & Methods

Four studies were undertaken with Paige Prostate Detect, an FDA-cleared AI-based algorithm that classifies digital whole slide images of prostate core needle biopsies as either 'suspicious' or 'not suspicious' for harboring cancer (Figure 2). Two studies used the algorithm as a standalone second read modality and considered the implications of a pre-screening use case through model calculations. The other studies employed a concurrent read design.

Literature Reference	Study Design	Time-Based Efficiency Endpoints
1 C Eloy, A. M. *et al* (2023). Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. *Virchows Archiv*, Epub Feb 2023	**Concurrent Read** • 4 general pathologists, N=41 cases containing N=105 biopsies • Unassisted and concurrent read arms with washout period	**Time required to review and report entire cases** *Comparison between concurrent read and unassisted diagnosis arms*
2 LM da Silva, E. P.-F. *et al* (2021). Independent real-world application of a clinical grade automated prostate cancer detection system. *Journal of Pathology, 254*(2)	**Second Read** as standalone modality • N=579 slides from N=100 patients	**Model calculation** comparing pre-screening scenario to unassisted diagnosis in terms of **total slide review time**
3 S Perincheri, A. L. *et al* (2021). An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Modern Pathology, 34*(8)	**Second Read** as standalone modality • N=1876 biopsies from N=118 patients	**Model calculation** comparing pre-screening scenario to unassisted diagnosis in terms of **total number of biopsies requiring review**
4 P Raciti, J. S. *et al* (2020). Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology, 33*(10)	**Concurrent Read** • 3 general pathologists, N=304 biopsies • Unassisted and Concurrent Read arms with washout period	**Time required to review single slide** *Comparison between Concurrent Read and unassisted diagnosis arms*

Table 1. Peer-reviewed literature analyzed where the efficiency of the market-authorized system Paige Prostate Detect was measured according to different reading paradigms

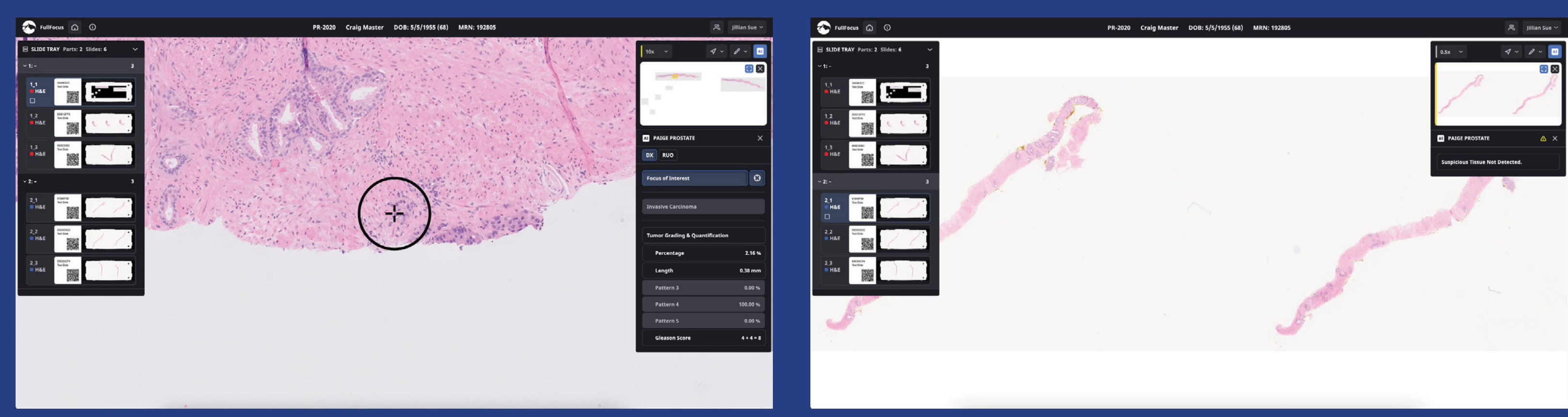

Figure 2. Example outputs of Paige Prostate Detect, AI-based software that classifies single whole slide images as either 'suspicious' (left) or 'not suspicious' (right) and presents outputs to users while reviewing tissue. Additionally, the outputs are available as binary slide classification per slide for non-current reading paradigms.

Results

Based on initial study outcomes and scenario modeling, the greatest time-based efficiency gains may be realized with a screening use case (65.5% reduction of slide review time[2]; 68.6% reduction of volume of slides reviewed[3]) or concurrent read (>20% timesaving[1,4]), while the additional step in a second read paradigm is unlikely to introduce time-based efficiency gains, Table 2. The literature also show that using an AI-based algorithm as a second read and concurrent read modality can significantly increase human+AI diagnostic accuracy, and potentially help pathologists identify otherwise missed foci of cancer.

A clear limitation of this analysis is that these data are comparing different studies on different data, and do not investigate the impacts of different reading paradigms on the same sample. Further, the pre-screening use case is not currently approved for clinical use and is modeled in the literature[2,3] though not implemented as a workflow. The pre-screening workflow is associated with a higher risk profile and regulatory burden, and is not the current intended clinical use of this system. Finally, there are important efficiencies to investigate beyond time—such as the reduction in the use of ancillary testing such as immunohistochemistry, and downstream impacts such as reduction in unnecessary patient interventions—that have been studied[1,5] and are more amenable to prospective studies involving live clinical deployments.

Reading Paradigm	Paige Prostate Detect: Comparisons to Unassisted Diagnosis	
	Accuracy Impacts	Time-Based Efficiency Impacts
Second Read	Increase in both sensitivity and specificity[2,3,5]	*No data currently available*
Concurrent Read	No statistically significant impact[1]; increase in sensitivity[4]	**Reduction of slide review time:** • 22% timesaving to review and report cases[1] • 21% timesaving to review cancerous slides[4]
Pre-Screening	*Not investigated*	65.5% reduction of total slide review time, modeled[3] 68.6% reduction of total slide volume to review, modeled[2]

Table 2. Summary of performance and efficiency impacts measured per reading paradigm using Paige Prostate Detect in the published literature.

Conclusions

In the EU and UK, health economics are a key consideration for adoption of AI algorithms and the screening and concurrent workflows would offer attractive efficiency gains over second read applications. Given the higher risk profiles of these use cases, a highly accurate and robust AI system is required, and further clinical validation guidelines should be developed for safe implementation in clinical practice.

References

1. C Eloy, A. M. *et al* (2023). Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. *Virchows Archiv,* Epub Feb 2023.
2. LM da Silva, E. P.-F. *et al* (2021). Independent real-world application of a clinical grade automated prostate cancer detection system. *Journal of Pathology,* 254(2).
3. S Perincheri, A. L. *et al* (2021). An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Modern Pathology*, 34(8).
4. P Raciti, J. S. *et al* (2020). Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology,* 33(10).
5. P Raciti and J Sue, J. A. R. *et al* (2022). Clinical Validation of Artificial Intelligence Augmented Pathology Diagnosis Demonstrates Significant Gains in Diagnostic Accuracy in Prostate Cancer Detection. *Archives of Pathology & Laboratory Medicine,* online ahead of print 2022.

Paige®