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. Retamer, MD, Jill Sue, MS, Brandon Rothrock, PhD, David S. Klimstra, MD
Paige.AI Inc., 11 Times Square, 37th 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 safety. Consequently, clinical performance studies are evolving from standard 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 refer to 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

1. Second Read
   - First, the pathologist reviews the images according to the standard of care and makes a diagnosis.
   - Then the pathologist views the results of the algorithm. This may or may not lead to revision of the original diagnosis.

2. Concurrent Read
   - In one session, the pathologist reviews the images and makes a diagnostic decision while having access to the results of the algorithm.
   - Second read (not shown).

3. Pre-screening
   - First, the algorithm outputs candidate areas likely to harbor malignancy in the images.
   - Then, the pathologist reviews only the areas that the algorithm has identified as ‘suspicions’ to make the final diagnosis.

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 ‘suspicous’ 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.

<table>
<thead>
<tr>
<th>Literature Reference</th>
<th>Study Design</th>
<th>Time-Based Efficiency Endpoints</th>
<th>Accuracy Impacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. C. Myllylä et al. (2020)</td>
<td>Concurrent Read</td>
<td>Increase in both specificity and sensitivity</td>
<td>No data currently available</td>
</tr>
<tr>
<td>2. N. D. Stamey et al. (2020)</td>
<td>Concurrent Read</td>
<td>No statistically significant impact in sensitivity</td>
<td>Reduction of false positive time: 22%</td>
</tr>
<tr>
<td>3. M. P. North et al. (2020)</td>
<td>Concurrent Read</td>
<td>No statistically significant impact in diagnostic accuracy</td>
<td>In terms of both false positive rate and time reduction</td>
</tr>
<tr>
<td>4. P. Roth et al. (2020)</td>
<td>Concurrent Read</td>
<td>Increase in both specificity and sensitivity</td>
<td>No data currently available</td>
</tr>
</tbody>
</table>

Table 1: Summary of performance and efficiency impacts associated with 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 risks of these use cases, a highly accurate and robust AI system is required, and further calibration validation guidelines should be developed for safe implementation in clinical practice.

References