Validating AI Apps for Pathology Practice

Liron Pantanowitz, MD
Professor of Pathology & Biomedical Informatics
Vice Chairman for Pathology Informatics
University of Pittsburgh Medical Center
pantanowitzl@upmc.edu
OUTLINE

1. AI and Anatomical Pathology
2. Questions for AI in Pathology
3. Validating AI in Clinical Practice
## WOMB TO TOMB

### Examples of AI applications across the human lifespan

<table>
<thead>
<tr>
<th>Embryo selection for IVF</th>
<th>Genomic interpretation</th>
<th>Sick newborns</th>
<th>Voice medical coach via a smart speaker (like Alexa)</th>
<th>K+</th>
<th>Mental health</th>
<th>Paramedic dx of heart attack, stroke</th>
<th>Assist reading of scans, slides, lesions</th>
<th>Prevent blindness</th>
<th>Classify cancer, identify mutations</th>
<th>Promote patient safety</th>
<th>Predict death in-hospital</th>
</tr>
</thead>
</table>

AI HYPE

Benefits
• Enhanced automation
• Eliminate tedious tasks
• More accurate diagnoses

Risks
• Biased algorithms
• Failed deployment
• Disruption
• Deskilled
• Replace humans
What worries you the most about AI in medicine? I have lots of worries. First, there’s the issue of privacy and security of the data. And I’m worried about whether the AI algorithms are always proved out with real patients. Finally, I’m worried about how AI might worsen some inequities. Algorithms are not biased, but the data we put into those algorithms, because they are chosen by humans, often are. But I don’t think these are insoluble problems.

Will we ever have an AI doctor to take care of all of our medical needs? The pinnacle of AI is being fully autonomous. But I don’t think that will happen in medicine; AI will always need human backup. A machine could handle certain things.
DIGITAL PATHOLOGY + AI

- Digital imaging (WSI) is mature (~20 years)
- Numerous vendors (hardware ± software)
- Multiple clinical & non-clinical applications
- Published literature validates the technology
- Regulatory approval (e.g. FDA) for clinical use
- Increasing interest but slow global adoption
- Digital Pathology is poised for an era of AI
- Vendor shift towards developing AI tools
- BUT still limited clinical experience with AI
PATHOLOGISTS SHOULD BE INVOLVED AT ALL STAGES OF AI
(DESIGN, DEVELOPMENT & VALIDATION)

MACHINE LEARNING

Digital image → Feature extraction → Classification → Output

[Diagram showing the process of machine learning]

DEEP LEARNING

Digital dataset → Convolutional Neural Network → Output

[Diagram showing the process of deep learning]
Questions for AI in Pathology


1. What are the right tasks for AI in Pathology?
2. What are the right data for developing AI tools?
3. What is the right evidence standard for using AI?
4. What is the right approach for using AI in practice?
WHAT ARE THE RIGHT DATA FOR AI?

• Both data *Quantity & Quality* are important
  • CNNs need large amount of good quality data
• Annotated & labeled data is necessary for ML
  • Expert annotation is expensive & a bottleneck
  • Crowdsourcing quicker & cheaper but has noise
  • Annotation problems (interface, ambiguity of features)
• WSI datasets required for training & validation
• 10,000 images is a suggested ballpark number
• For limited datasets try augmentation, etc.
WORKING WITH INDUSTRY

• Academic Industry Partnerships
• Key part of “innovation cycle”
• Commercialization of WSI
• Rules of engagement:
  • Conflict of interest
  • Intellectual property
  • Transparency
  • Sunshine Law
  • Professionalism
• Ask for permission before forgiveness!
WHAT IS THE RIGHT EVIDENCE STANDARD FOR AI?

• Peer reviewed scientific publications
  • Algorithm performance (accuracy)

• Clinical validation (trials)
  • Demonstrate improved outcomes
  • Lack of unintended consequences

• Business use case
  • Cost-effectiveness of AI

• FDA approval
  • Re-assurance of app safety
  • Less burden for self-validation
Most of the algorithms have been developed on a research basis or in a test environment, and only recently applied. I've been working with one or two of these image algorithms that provide AI tools in order to do the following:

1. **Validate these algorithms on our own images.** Although we produce an H&E stain, it's probably unique to our environment, so it would be important to make sure that it would work in an AI app use at UPMC.

2. **Explore possibilities to embed these AI algorithms into clinical work flow.** Assuming that an algorithm has been proven to work and is feasible, there is the issue of how exactly to embed these apps into our clinical workflow. Where do you insert it into your workflow and what kind of infrastructure is required in our environment for this to work?
INTEGRATING AI INTO CLINICAL PRACTICE

• Practical for operational use
  • Minimize workflow disruption
  • Interoperability (LIS-centric)
  • Scalable and parallelizable
  • Manageable IT needs
  • Explainable AI & human readable

• Proper oversight
• ROI for using AI (reimbursement)
• Change management
  • MD vs Machine mindset
HOW TO INTEGRATE AI TOOLS IN CLINICAL PRACTICE

Scanning

Primary Diagnosis

Al to QA check now & later

Al to QA, screen, pre-diagnose & pre-order stains

Al consultant when directed
AI & THE CLOUD

Cloud-Based App
## IT INFRASTRUCTURE REQUIRED

<table>
<thead>
<tr>
<th>IT ISSUE</th>
<th>CURRENT CONCERNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• GPUs instead of CPUs</td>
<td>Are these available in house?</td>
</tr>
<tr>
<td>• On-Premise vs Cloud</td>
<td>Which platform works better?</td>
</tr>
<tr>
<td></td>
<td>(implementation, customization, performance, maintenance)</td>
</tr>
<tr>
<td></td>
<td>Which deployment costs less?</td>
</tr>
<tr>
<td></td>
<td>(Capital vs operating expenditure)</td>
</tr>
<tr>
<td></td>
<td>What are the security concerns?</td>
</tr>
<tr>
<td></td>
<td>Who owns our lab’s stored data?</td>
</tr>
<tr>
<td>• Workflow and reporting</td>
<td>Is LIS integration required?</td>
</tr>
<tr>
<td></td>
<td>(triage with reflex testing &amp; CMS rules, reporting)</td>
</tr>
</tbody>
</table>
EXPLAINABLE AI

Heat Maps


Gallery Display

Portion of H&E image with algorithm results (red = high probability for cancer)
Impact of Data Presentation on Physician Performance Utilizing Artificial Intelligence-Based Computer-Aided Diagnosis and Decision Support Systems

L. Barinov 1, 2, 3, A. Jairaj 1, M. Becker 3, 4, S. Seymour 1, E. Lee 3, 4, A. Schram 3, 4, E. Lane 4, A. Goldszal 3, 4, D. Quigley 4, L. Paster 3, 4

<table>
<thead>
<tr>
<th>Radiologist ID</th>
<th>Sequential read AUC, 95% CI</th>
<th>P value CR vs SR two-tailed alpha = 0.05</th>
<th>Independent read AUC, 95% CI</th>
<th>P value CR vs IR two-tailed alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7935 [0.7567–0.8229]</td>
<td>0.235</td>
<td>0.8213 [0.7861–0.8516]</td>
<td>0.0285*</td>
</tr>
<tr>
<td>2</td>
<td>0.7674 [0.7327–0.8001]</td>
<td>0.601</td>
<td>0.8305 [0.7982–0.8594]</td>
<td>0.00155*</td>
</tr>
<tr>
<td>3</td>
<td>0.7859 [0.7527–0.8174]</td>
<td>0.0532</td>
<td>0.7988 [0.7632–0.8310]</td>
<td>0.0160*</td>
</tr>
</tbody>
</table>

*Significant

- Using AI-based decision support used in US image analysis
- 2 different reading methods: sequential & independent
- Significant differences with accuracy & inter-operator variability
- Practical implications for utilization of AI in diagnostic environments.
BREAST CARCINOMA WITH MITOTIC FIGURES
ARE ALGORITHMS GENERALIZABLE?

• Heterogeneity (e.g. H&E stain variation)
• Problem with limited data sets (no diversity, algorithmic bias)
• Responsibility of labs to thus calibrate apps prior to use
• Does this need to be occur periodically to catch “drift”?
### Example Clinical Validation

<table>
<thead>
<tr>
<th>Phase</th>
<th>Calibration</th>
<th>Analytical Validation</th>
<th>Compliance Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td># Slides (WSIs)</td>
<td>30 (300+)</td>
<td>100 (2000+)</td>
<td>60 (120+)</td>
</tr>
<tr>
<td>Rationale</td>
<td>Check that the AI tool works as specified (vs a reference)</td>
<td>Shows the AI tool consistently performs as expected</td>
<td>Confirms AI tool is used safely and compliant</td>
</tr>
</tbody>
</table>
TAKE-HOME MESSAGE

• There is limited clinical experience with AI to know how to best use this technology in clinical practice

• AI for clinical diagnostic work needs to be practical, interoperable, explainable, generalizable and manageable

• Take care how AI is deployed because the way in which it is used can have unwanted consequences

• Recommendations &/or guidelines are required to promote the safe use & adoption of AI in our labs