Machine learning for Pathology

Andrew H Beck MD PhD
CEO @ PathAI

March 19, 2019
Pathologic diagnosis is a central determinant of therapeutic decisions.
Immunohistochemistry – reference these images

http://www.pathologyoutlines.com/stains.html
Emergence of early computational approaches in Pathology (1981)

MORPHOMETRY FOR PROGNOSIS PREDICTION IN BREAST CANCER

Sir,—Some workers have found a correlation between prognosis and microscopical features of the primary tumour in breast cancer\(^1\)–\(^3\) but in one large prospective study the significance of the nuclear and histological grade for prognosis was weak.\(^4\) Disagreement in grades assigned to the same tumours by different pathologists may range up to 40%\(^5\),\(^6\) and this disagreement may be due to the subjective nature of histopathological assessment. In contrast, the advantages of morphometry are objectivity and high reproducibility.\(^7\)

<table>
<thead>
<tr>
<th>Method</th>
<th>Total (n = 78)</th>
<th>Learning set (n = 38)</th>
<th>Test set (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANS</td>
<td>59</td>
<td>65</td>
<td>54</td>
</tr>
<tr>
<td>TNM</td>
<td>64</td>
<td>67</td>
<td>56</td>
</tr>
<tr>
<td>Morphometry</td>
<td>87</td>
<td>92</td>
<td>78</td>
</tr>
</tbody>
</table>

Baak et al. Lancet 1981
Artificial Neural Nets in Quantitative Pathology (1990)

“It is concluded that artificial neural networks, used in conjunction with other nonalgorithmic artificial intelligence techniques and traditional algorithmic processing, may provide useful software engineering tools for the development of systems in quantitative pathology.”
Emergence of Digital Pathology (2000)


Digital Pathology: Science Fiction?

Mattia Barbareschi,* Francesca Demichelis,† Stefano Forti,†
and Paolo Dalla Palma*

But what will come next? Is it possible to hypothesize that VC will completely substitute our traditional glass slides? Maybe yes, and let us describe the “science fiction” new millennium digital pathology laboratory, which we will call “DIGIPATH.”
The emergence of machine learning-based approaches for cancer histopathology

Beck ... Koller. Science Translational Medicine 2011
Extracting a rich quantitative feature set

c. Constructing higher-level contextual/relational features:

- relationships of contiguous epithelial regions with underlying nuclear objects
- relationships between epithelial nuclear neighbors
- relationships between morphologically regular and irregular nuclei
- relationships between epithelial and stromal objects
- relationships between epithelial nuclei and cytoplasm
- characteristics of epithelial nuclei and epithelial cytoplasm
- characteristics of stromal nuclei and stromal matrix

Beck ... Koller. Science Translational Medicine 2011
C-Path 5YS Score Significantly Associated with Overall Survival on Both Cohorts

Beck ... Koller. Science Translational Medicine 2011
Even today, the anatomic path lab has been largely unchanged for routine diagnostics.
And core technology breakthroughs in routine use are from the 19th century.

Histochemical Stains
Developed from combinations of analine and natural dyes in the later half of the 19th century.

Photomicroscope
Horizontal apparatus with camera, microscope, and light source, 1895.
Discordance among pathologists is common in interpretation of breast biopsies

Pathologists in individual practice setting

Overall concordance rate of 75% on breast biopsies.

Inter-observer concordance rate of only 48% for a diagnosis of atypia.

Intra-observer concordance is only 79% overall and 53% for atypical lesions

Ref: Jackson SL ... Elmore JG. Ann Surg Oncol. 2017 May;24(5):1234-1241.
Discordance among pathologists is common in interpretation of melanocytic neoplasms on skin biopsies

- 187 pathologists interpreted skin lesion biopsies, resulting in an overall discordance of 45%
- 118 pathologists read the same samples 8 months apart, and had an intraobserver discordance of 33%
Discordance rates across a broad set of specimen types is fairly high with little improvement over past several decades.

Table 3. Summary of Studies That Express a Discrepancy or Major Discrepancy Rate

<table>
<thead>
<tr>
<th>Study Type</th>
<th>No. of Studies</th>
<th>Median (25th–75th Percentile)</th>
<th>No. of Studies</th>
<th>Median (25th–75th Percentile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td>116&lt;sup&gt;c&lt;/sup&gt;</td>
<td>18.3 (7.5–34.5)</td>
<td>78&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5.9 (2.1–10.5)</td>
</tr>
<tr>
<td>Surgical pathology</td>
<td>84&lt;sup&gt;e&lt;/sup&gt;</td>
<td>18.3 (7.5–37.4)</td>
<td>63&lt;sup&gt;f&lt;/sup&gt;</td>
<td>6.3 (1.9–10.6)</td>
</tr>
<tr>
<td>Cytology</td>
<td></td>
<td>24.8 (17.4–38.8)</td>
<td>11&lt;sup&gt;h&lt;/sup&gt;</td>
<td>4.3 (2.8–7.5)</td>
</tr>
<tr>
<td>Both</td>
<td></td>
<td>9.1 (6.7–15.8)</td>
<td>11&lt;sup&gt;i&lt;/sup&gt;</td>
<td>5.9 (3.3–8.7)</td>
</tr>
<tr>
<td>Multigorgan</td>
<td></td>
<td>9.1 (3.8–18.7)</td>
<td>42&lt;sup&gt;l&lt;/sup&gt;</td>
<td>3.9 (1.1–7.4)</td>
</tr>
<tr>
<td>Single-organ&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>25.2 (14.0–43.7)</td>
<td>36&lt;sup&gt;n&lt;/sup&gt;</td>
<td>8.0 (3.7–15.8)</td>
</tr>
<tr>
<td>Internal&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>10.9 (3.8–17.6)</td>
<td>22&lt;sup&gt;p&lt;/sup&gt;</td>
<td>1.2 (0.30–3.1)</td>
</tr>
<tr>
<td>External</td>
<td>79&lt;sup&gt;q&lt;/sup&gt;</td>
<td>23.0 (10.6–40.2)</td>
<td>56&lt;sup&gt;r&lt;/sup&gt;</td>
<td>7.4 (4.6–14.7)</td>
</tr>
</tbody>
</table>

High Error Rates

Massive advances in deep learning for computer vision...
What does AI mean at PathAI?

• Models which learn how to make decisions and predictions by recognizing patterns in data.
• These can be traditional machine learning models or, more commonly, deep convolutional neural networks.

The human defines the data, the data defines the algorithm.

Traditionally, the human defines the algorithm.
What can AI do for pathology?

**A (somewhat) practical treatment**

- Exhaustive – the model is tireless and is not distracted
- Quantitative – the model is reproducible and objective
- Efficient – massive parallelization for speedy processing
- Exploratory - learn relationships in a purely data-driven manner
What AI can’t do for pathology

Replace pathologists!
A diagnosis/detection example:

Breast cancer metastases

• After a primary mass discovered, lymph nodes are biopsied
• Pathologists check these for metastases
• Non-zero failure rate: a retrospective study found a 24% disagreement rate

The data - CAMELYON

- H & E stained, Formalin-Fixed, Paraffin-Embedded (FFPE)
  - 270 training slides, 129 test
- Annotated by a panel
The data – Whole-Slide Images

- WSIs are large -- 20,000-200,000 pixels on a side ("gigapixel")
  - mm-cm imaged at 20x/40x
- Demo – TCGA lung cancer
Approach

• Standard image classification approach needs a twist for WSIs: sampling
Successfully applied deep learning approach to pathology

Our team won the Camelyon challenge in 2016, demonstrating outstanding initial performance in pathology.

Deep learning model outperforms human pathologists in the diagnosis of metastatic cancer

<table>
<thead>
<tr>
<th></th>
<th>Error Rate (1-AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologists in competition</td>
<td>3.5%</td>
</tr>
<tr>
<td>Pathologists in clinical practice(^1)</td>
<td>13 – 26%</td>
</tr>
<tr>
<td>Pathologists on micro-metastasis(^2)</td>
<td>23 – 42%</td>
</tr>
<tr>
<td>Deep learning model</td>
<td>0.65%</td>
</tr>
</tbody>
</table>

\(^1\) n=12  \(^2\) Small tumors

Caveats and considerations

- Real world data vs. competition data

Figure 2. Computer plays *Name That Dataset*. Left: classification performance as a function of dataset size (log scale) for different descriptors (notice that performance does not appear to saturate). Right: confusion matrix.

Torralba & Efros, 2011
Pathologist + PathAI
**Pathology Report**

<table>
<thead>
<tr>
<th>Patient: John Doe</th>
<th>pTNM staging:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis: Met. Cancer</td>
<td>pT2N1MX</td>
</tr>
<tr>
<td>Size: 2.3mm</td>
<td># of Pos LN: 1</td>
</tr>
<tr>
<td># of Neg LN: 4</td>
<td></td>
</tr>
<tr>
<td>Time per slide: 10 – 60 seconds</td>
<td>Accuracy: &gt;99.5%</td>
</tr>
<tr>
<td>Reproducibility: High</td>
<td></td>
</tr>
</tbody>
</table>
Why is this a good application for AI?

- Exhaustive analysis is beneficial
  - Large volume
- Local image data necessary and sufficient
- Interpretability: Heatmaps & simple models provide insight into how the patient-level prediction was made
- Required accuracy is high
A predictive example: Precision immunotherapy

• Some cancers express immune-inhibitory ligands, activating immune “checkpoints”

• “checkpoint inhibitors” mask these signals, unleashing the immune system
A predictive example: Precision immunotherapy

- Response rate is low, but some fraction of patients are essentially “cured”
- PD-L1 expression is somewhat indicative of response
Manual interpretation of PD-L1 IHC is highly variable

**PDL1 manual IHC scores on immune cells are unreliable**

Table 2. ICC for the Pathologist Scores and Concordance Statistics

<table>
<thead>
<tr>
<th>Cells</th>
<th>Antibody, ICC (95% CI)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Summary, Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22c3</td>
<td>28-8</td>
<td>SP142</td>
<td>E1L3N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor cells</td>
<td>0.882 (0.873-0.891)</td>
<td>0.832 (0.820-0.844)</td>
<td>0.869 (0.859-0.879)</td>
<td>0.859 (0.849-0.869)</td>
<td>0.86 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Immune cells</td>
<td>0.207 (0.190-0.226)</td>
<td>0.172 (0.156-0.189)</td>
<td>0.185 (0.169-0.203)</td>
<td>0.229 (0.211-0.248)</td>
<td>0.19 (0.03)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: ICC, intraclass correlation coefficient.

aN = 90.
Manual scoring of PD-L1 is variable ...and not always predictive

RESULTS
The median overall survival was 9.2 months (95% confidence interval [CI], 7.3 to 13.3) with nivolumab versus 6.0 months (95% CI, 5.1 to 7.3) with docetaxel. The risk of death was 41% lower with nivolumab than with docetaxel (hazard ratio, 0.59; 95% CI, 0.44 to 0.79; \( P<0.001 \)). At 1 year, the overall survival rate was 42% (95% CI, 34 to 50) with nivolumab versus 24% (95% CI, 17 to 31) with docetaxel. The response rate was 30% with nivolumab versus 9% with docetaxel (\( P=0.008 \); 0.47 to 0.81; \( P<0.001 \)). The expression of the PD-1 ligand (PD-L1) was neither prognostic nor predictive of benefit. Treatment-related adverse events of grade 3 or 4 were reported in 7% of the patients in the nivolumab group as compared with 55% of those in the docetaxel group.
Can we do better?

- Deep learning is data hungry
- Need 10s of thousands of precise cell annotations

First, we need the data
Board-certified training data

Working with pathologists around the country to generate high-quality annotations

Total annotations, 2017 - 2019

>2.5M Annotations
Automatic and exhaustive regions of interest
tumor and relevant stroma
IHC expression difficult to detect on immune and tumor cells
Exhaustive automated classification
Cell type and cellular IHC positivity classification
Quantitative and reproducible PD-L1 scoring

- Manual review: few hundred cells over a few arbitrary high-power fields of view
- Automated analysis: exhaustive classification of 10k-1million cells
Taking it further
From quantitative assay to patient prediction

- PD-L1 scoring alone reduces billions of pixels to 1-2 numbers.
- Can we identify additional relevant information?
  - Using data from randomized controlled clinical trials

- However: Millions of patches, *hundreds* of patients
Predictive features guided by biomedical priors

H & E slide matching PD-L1 slide
Predictive features guided by biomedical priors
Immune cell (lymphocyte) detection
Predictive features guided by biomedical priors
Cancer epithelium (red) and stroma (green) segmentation
Predictive features guided by biomedical priors

Epithelial-stromal interface definition
Cell-type specific, tissue context-aware IHC-quantification
Data-driven identification of pathological phenotypes associated with drug response

- Total number of macrophages in epithelial/stroma interface (80um)
- Total number of macrophages in epithelial/stroma interface (120um)
- Total number of macrophages in invasive margin (250um)
- Total number of lymphocytes in epithelial/stromal interface on H&E stain
- Total number of plasma cells in epithelium on H&E stain
- Total number of plasma cells in stroma on H&E stain
- Tumor (epithelium + stroma) area on H&E stain
- Total number of plasma cells in epithelial/stroma interface (40um)
- Total number of plasma cells in epithelial/stroma interface (80um)
- Area (mm$^2$) of epithelial/stroma interface (80um) target positive cancer cells on target stain
- Area (mm$^2$) of epithelial PDL-1 positive macrophages on target stain
- Necrosis area on target stain
- Proportion of tumor infiltrating lymphocytes engaged by target positive macrophages
- Stroma area on target stain
- Tissue area on target stain
Multivariate models predictive of IO response

• Low \( n \), interpretability and measures of uncertainty valuable:
  • No deep learning (gasp!)
• Feature importance/selection in these models can provide disease insight
  • Now we’re doing things pathologists can’t rather than automating / improving what they already can

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Hazard ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>161</td>
<td>0.89 (0.60, 1.32)</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Note: KM curves for illustration only
How do we know these features are correct?

Frames
Validation by exhaustive consensus
Many other application areas
The Cancer Genome Atlas - Melanoma

TCGA-EE-A2GL, Malignant Melanoma
Melanoma Tissue Map

Tumor Stroma

Tumor Epithelium

TCGA-EE-A2GL, Malignant Melanoma
Melanoma Cell Map

Lymphocytes: Green
Macrophages: Orange
Plasma Cells: Blue
Fibroblasts: Yellow
Melanoma Cells: Red
Exhaustive analysis of cellular features in TCGA to enable data-driven identification of pathological predictors of survival in malignant melanoma

Pathological phenotypes with FDR < 5% for association with Progression Free Survival

Increased area of stromal plasma cells associated with improved survival in melanoma
Data-driven identification of transcriptional signature underlying stromal area of plasma cells in melanoma

Top-ranking transcripts associated with stromal area of plasma cells

<table>
<thead>
<tr>
<th>Gene</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>REC8</td>
<td>0.57</td>
</tr>
<tr>
<td>GPR174</td>
<td>0.53</td>
</tr>
<tr>
<td>CD38</td>
<td>0.53</td>
</tr>
<tr>
<td>LAX1</td>
<td>0.53</td>
</tr>
<tr>
<td>TOX</td>
<td>0.53</td>
</tr>
<tr>
<td>AKAP5</td>
<td>0.53</td>
</tr>
<tr>
<td>C8orf80</td>
<td>0.52</td>
</tr>
<tr>
<td>JSRP1</td>
<td>0.52</td>
</tr>
<tr>
<td>IGJ</td>
<td>0.52</td>
</tr>
<tr>
<td>TNFRSF17</td>
<td>0.51</td>
</tr>
<tr>
<td>EAF2</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Stromal plasma cell area RNA signature strongly enriched for immune genes

<table>
<thead>
<tr>
<th>Gene Set Name</th>
<th>Description</th>
<th>FDR q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>REACTOME_IMMUNE_SYSTEM</td>
<td>Genes involved in Immune System</td>
<td>7.62E-57</td>
</tr>
<tr>
<td>REACTOME_ADAPTIVE_IMMUNE_SYSTEM</td>
<td>Genes involved in Adaptive Immune System</td>
<td>6.02E-42</td>
</tr>
<tr>
<td>PID_TCR_PATHWAY</td>
<td>TCR signaling in naive CD4+ T cells</td>
<td>4.24E-30</td>
</tr>
<tr>
<td>REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL</td>
<td>Genes involved in Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell</td>
<td>6.07E-26</td>
</tr>
<tr>
<td>KEGG_PRIMARY_IMMUNODEFICIENCY</td>
<td>Primary immunodeficiency</td>
<td>7.98E-24</td>
</tr>
<tr>
<td>PID_IL12_2PATHWAY</td>
<td>IL12-mediated signaling events</td>
<td>9.27E-24</td>
</tr>
<tr>
<td>PID_CD8_TCR_PATHWAY</td>
<td>TCR signaling in naive CD8+ T cells</td>
<td>9.27E-24</td>
</tr>
<tr>
<td>KEGG_CELL_ADHESION_MOLECULES_CAMS</td>
<td>Cell adhesion molecules (CAMs)</td>
<td>3.00E-22</td>
</tr>
<tr>
<td>KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION</td>
<td>Cytokine-cytokine receptor interaction</td>
<td>6.38E-22</td>
</tr>
<tr>
<td>KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION</td>
<td>Intestinal immune network for IgA production</td>
<td>3.37E-21</td>
</tr>
<tr>
<td>REACTOME_TCR_SIGNALING</td>
<td>Genes involved in TCR signaling</td>
<td>3.24E-20</td>
</tr>
<tr>
<td>REACTOME_PD1_SIGNALING</td>
<td>Genes involved in PD-1 signaling</td>
<td>3.44E-19</td>
</tr>
<tr>
<td>REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY</td>
<td>Genes involved in Costimulation by the CD28 family</td>
<td>5.48E-19</td>
</tr>
</tbody>
</table>
Another AI plus: scalability

- Same pipeline for any solid tumor type
- Contrast to traditional approach: hand-crafted algorithms.

PathAI for Immuno-oncology

PathAI platform has been applied to:
- Non-small cell lung cancer (Adenocarcinoma)
- Non-small cell lung cancer (Squamous Cell Carcinoma)
- Small cell carcinoma of the lung
- Urothelial carcinoma of the bladder
- Head and neck squamous cell carcinoma
- Melanoma
- Breast cancer
- Prostate cancer
- Colon cancer

>30 IO-IHC biomarkers studied

IHC images processed
10,000+

Number of Annotations
2.5 Million+

PDIL IHC cells classified
1 Billion+

In 2018, PathAI classified ~15x the number of cells that all US pathologists could perform in a year.
Extensive Slide Search & Data Standardization

Slides Search

Filter Images

Choose criteria

TCGA

TCGA

Any case

Any stain

Any group

Original file name

Overlays:

Yes  No  Either

Annotations:

Yes  No  Either

30872 matching images  Clear filters
Automated quality control

Blurred areas

Debris

Folded / damaged tissue
Annotate, train and deploy task-specific models

- Determined by partner needs
Interpretable feature extraction

- Hypothesis & data driven
Interactive Reports & Live Project Progress

**Melanoma Study Project**

**Overview**

The goal of this project is to leverage the PathAI platform to quantitate cellular and morphologic phenotypes from IHC (PD-L1) stained images in melanoma clinical trial data sets. The algorithms developed will be validated using exhaustive annotations on selected window frames, and algorithm improvements will be implemented to include new features and rule-based region-of-interest (ROI) selection. Once validated, extracted image features will be used to find associations with patient clinical outcomes (Best OR, PFS, OS).

**Progress**

- PathAI added a key result. 2d ago
- Project status changed to Predictive analysis. 3d ago
- PathAI uploaded a report. 3d ago
- PathAI released slide overlays Cell Detection v1, Tissue map v1 21d ago
- Project status changed to Predictive analysis.
The PathAI Deep Learning Process

Whole-Slide Images + Data
Transmit training data securely to the PathAI cloud

Annotations
Network of board-certified pathologists to provide ground truth consensus

Deep Learning Analysis
Cell detection, tissue & region classification

Deep Learning Feature Analysis
Over 200 relevant features extracted, measured and analyzed

Assay Validated
Identified features of significance reduced to practice

Assay Deployed
Analyze samples, quantified & visual results delivered

We can execute process in 4 – 8 weeks for new assays
AI in medicine

Some closing thoughts

- ML in the real world:
  - Building the right dataset is 75% of the challenge
- Modern ML: engineering and empirical science
  - Rigorous validation is key
- Ideas and algorithms vs. teams and infrastructure
Core challenges and road ahead

Technology
Regulatory
Financial
Workflow transformation
Key Takeaways

• Researchers have been working on AI for pathology for ~30 years

• In the past 5 years, advances in:
  • Availability of digital data
  • Access to large-scale computing resources
  • Major algorithmic advances (e.g., Deep CNNs)

• AI works extremely well when these 3 factors are all available and fails when they are not
Key Takeaways

• AI-powered pathology is broadly applicable across all image-based tasks in pathology and enables integration with other structured data types (e.g., ‘Omics)

• As AI and digital pathology are incorporated into clinical workflow, they will offer significant operational and efficiency advantages

• AI will drive improvements in the accuracy and predictiveness of pathology leading to research advances and improved care for patients
“In the Future...” (1987)

• “Integrated information systems, patient care management by exception, decision support tools, and, in the future, "artificial intelligence" assists can all be expected to become staples of pathology practice, especially impacting those pathologists who choose to be responsive to the new practice milieu of medical information science.”
Thank you!

The PathAI team

pathai.com
Opportunities for ML engineers/scientists, software engineers, pathologists,...