A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD
Agenda

• Brief comments
  – Introduce an interesting reference (Eric Topol, *Deep Medicine*)

• Introduce the speakers
  – Anand S. Dighe, MD, PhD (Massachusetts General Hospital)
  – Michael D. Feldman, MD, PhD (University of Pennsylvania)
  – Andrew R. Janowczyk, PhD (Case Western Reserve University)
Will artificial intelligence replace doctors?

Several new studies have shown that computers can outperform doctors in cancer screenings and disease diagnoses. What does that mean for newly trained radiologists and pathologists?
The outlandish expectations for AI in healthcare, a partial list (Table 1.1 in Deep Medicine by Eric Topol)

- Outperform doctors at all tasks
- Diagnose the undiagnoseable
- Treat the untreatable
- See the unseeable on scans, slides
- Classify the unclassifiable
- Eliminate workflow inefficiencies
- Eliminate hospital admissions and readmissions
- Eliminate the surfeit of unnecessary jobs
- 100% medication adherence
- Zero patient harm
- Cure cancer

© 2019 College of American Pathologists. Materials used with permission of faculty.
# Table 1.2 from Deep Medicine: changes in Medicine from 1975 to now

<table>
<thead>
<tr>
<th>Metric</th>
<th>1975</th>
<th>Now</th>
</tr>
</thead>
<tbody>
<tr>
<td># healthcare (HC) jobs</td>
<td>4 Million</td>
<td>&gt;16 Million</td>
</tr>
<tr>
<td>HC spending per person per year</td>
<td>$550</td>
<td>$&gt;11,000</td>
</tr>
<tr>
<td>Time allocated for office visits</td>
<td>60 min. new</td>
<td>12 min. new</td>
</tr>
<tr>
<td></td>
<td>30 min. return</td>
<td>7 min return</td>
</tr>
<tr>
<td>% GDP for HC</td>
<td>&lt; 8</td>
<td>18</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>None of these</td>
<td>RVUs, EHRs, PBMs, “health systems”</td>
</tr>
</tbody>
</table>
Future of AI (from Deep Medicine)

- Deep phenotyping (Ability to deeply define each individual using all relevant data (medical, social, behavioral, family history, anatomy, physiology and environment))
- Deep learning (wide range of conditions)
- Deep empathy and connection – greatest opportunity
Where AI lives the hype

• Specific pattern based recognition in specific fields
  • Needs careful training and validation (more coming)
A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD
Clinical Laboratory Informatics: New Tools to Improve Quality and Enhance Value

Director, MGH Core Laboratory
Director of Clinical Informatics
Clinical Lead, Partners Enterprise Pathology and eCare
Associate Professor, Harvard Medical School
Massachusetts General Hospital
Boston, MA
Role of informatics in the redesign of two key pathology/healthcare processes:

1. Test ordering
2. Result generation/interpretation
"The Dark Ages" (1990s)

<table>
<thead>
<tr>
<th>IMMUNOLOGY SECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>R x SPEP + Quant. IgG, IgA, IgM (IEPAP) R Free κ/λ Light Chains w/ Ratio (EKLBB)</td>
</tr>
<tr>
<td>R IgG R IgA R IgM R IgE R Viscosity R H. pylori IgG Ab R Haphtoglob.</td>
</tr>
<tr>
<td>R x Antinuclear Ab R x RF (RHP) R a CQ P R x Hyper. Pneu. Panel (HPS) R x a 1-AT</td>
</tr>
<tr>
<td>R x aDNA R x a 0/40 R x aSm R x RNP R x a Jo1 (JO1) R x a ScI70 R x Ceruloplasmin</td>
</tr>
<tr>
<td>R Mitochondrial Ab (AMA) R Smooth Muscle Ab (SMA) R Gastric Parietal Cell Ab (AGPC)</td>
</tr>
<tr>
<td>u x EJ protein (UJBP) u x Urine Free κ/λ w/ Ratio (UCLRB) □ Crystals, joint fluid (JFCP)</td>
</tr>
</tbody>
</table>

Send on Ice:
R □ Total Complement (CATE) □ C3 □ C4
R □ C1 inhibitor protein □ Factor B (PFB)

Send warm, 37°C (Separate Requisition & Bag Required)
2R □ Cryocrit only (CRYCRT)
2R □ Cryocrit + identification (CRY)

<table>
<thead>
<tr>
<th>DIABETES SECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>L □ Hgb A1C (A1CC) □ by affinity, circle reason: sickle cell thalassemia HgbF other:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GI SECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>R □ Ab to Hep A (IgG / IgM) (ANTHAV)</td>
</tr>
<tr>
<td>R □ Ab to Hep B Core Ag, Total (ANTHBC)</td>
</tr>
<tr>
<td>R □ Hep B e Ag and Ab (HBEAA)</td>
</tr>
<tr>
<td>P □ Hep B Viral DNA (HBVQ)</td>
</tr>
<tr>
<td>R □ Ab to Hep Delta Virus (ANTHDV)</td>
</tr>
<tr>
<td>P □ Hep C Viral RNA Quant</td>
</tr>
<tr>
<td>P □ Hep C Viral RNA Qual (HCVQL)</td>
</tr>
<tr>
<td>P □ Hep C Viral Genotype (HCVGN)</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.
Permits Pathology to have control over Provider Order Entry screens

- Synchronizes with Laboratory Information System (LIS)
- Pathology control of ordering messages, alerts, search terms, related tests
- Allows cataloging of Pathology data such that it can be shared via web services

Providers

Inpatient/ED/Outpatient Provider Order Entry

MGH PathConnect Middleware

Lab Orders

Laboratory Information System

Laboratory Staff

Am J Clin Pathol. 2010;133:860

© 2019 College of American Pathologists. Materials used with permission of faculty.
After extensive evaluation from Partners selected Epic as the vendor for the enterprise-wide implementation of Partners eCare

Total cost of the 7 year implementation (2013-2019) across Partners was $1.9 billion dollars
The “Good Old Days” (2000-2015: homegrown custom lab software)

Permits Pathology to have control over Provider Order Entry screens

- Synchronizes with Laboratory Information System (LIS)
- Pathology control of ordering messages, alerts, search terms, related tests
- Allows cataloging of Pathology data such that it can be shared via web services

© 2019 College of American Pathologists. Materials used with permission of faculty.
“If you're yearning for the good old days, just turn off the air conditioning.”

- Griff Niblack

The “Good Old Days:”

- Limited ability for Pathology to intervene in care pathways
- Limited ability for Pathology to interact with provider orders
- Limited ability to assess outcomes
- Poor infrastructure in place for decision support
- Challenges in obtaining data for projects
- Computational techniques not widely used in Pathology
Enterprise Lab Information System Governance

Pre-Epic
- Local LIS teams at each hospital (large teams at MGH and BWH)

Post-Epic
- **Single enterprise LIS team** for all enterprise lab functions
- Enterprise LIS team **tightly integrated with the Epic orders team**
- **Shared tools, org chart, and infrastructure**
- **EHR enterprise lab leadership structure**

*All IT and LIS teams co-localized in new facility*
Pathology Datamart (Pre-EHR)

Input

Laboratory Information System

Pathology Datamart

Daily result reporting detail
Pathology Datamart (Post-EHR)

Inputs

- Daily reports of orders and result messages
- Expanded test profiles
- Daily result reporting detail

Pathology Datamart

Outputs

- Lab dashboard
- Online lab handbook
- Key Epic/Lab reports

© 2019 College of American Pathologists. Materials used with permission of faculty.
Epic Reporting Tools

Daily reports available of all Epic orders including problem list, diagnosis, order source (order set or preference list)

Daily monitoring needed to assess for utilization issues

- Quick, real time data
- Little flexibility in customization
- Actionable to access patient record

- Real time & small data set
- More flexibility in customization
- Actionable to patient record

Clarity
- Refreshed every night
- Analytical trending data
- Highly customizable

© 2019 College of American Pathologists. Materials used with permission of faculty.
Laboratory Metrics: Online Dashboard

Used pathology and Epic data to create a real time dashboard to measure Epic and Laboratory outcomes and permit rapid course corrections.

Outpatient Turnaround Time

**Pre-Epic**: Collect to Result = 237 min

**Post-Epic**: Collect to Result = 143 min
## Adapting to Enterprise Information Systems

<table>
<thead>
<tr>
<th></th>
<th>Homegrown system</th>
<th>Enterprise system (Epic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menu size</td>
<td>Limited (95%)</td>
<td>Most tests available (99%)</td>
</tr>
<tr>
<td>In lab processing</td>
<td>Manual steps, slow</td>
<td>Rapid, efficient</td>
</tr>
<tr>
<td>Lab test search</td>
<td>Provides decision support, CDS visible when searching</td>
<td>Search capabilities primitive, does not store search results or provide visible CDS when searching</td>
</tr>
<tr>
<td>Ordering favorites</td>
<td>Not permitted</td>
<td>Allowed</td>
</tr>
<tr>
<td>Order sets</td>
<td>Reviewed by lab</td>
<td>Uncommonly reviewed by lab</td>
</tr>
<tr>
<td>Collection process</td>
<td>Simple (but manual)</td>
<td>Complex (but electronic)</td>
</tr>
<tr>
<td>Decision support availability</td>
<td>Custom, fast, not requiring programming</td>
<td>Extensive possibilities but requires many levels of approvals, implementation complex</td>
</tr>
<tr>
<td>Utilization control</td>
<td>Able to easily manipulate menu, CDS to influence testing</td>
<td>More challenging</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.
New Responsibilities for Pathologists in the EHR Era

### Structural
- Menu, synonyms, search
- Order configurations
- Frequency restrictions
- Order sets/panels

### Collection
- Order release times
- Cancellation rules
- Combining rules

### Result Harmonization
- Point of Care results
- External results
- EHR to EHR results

### Results Routing
- EHR Inbox monitoring
- Post-discharge result routing

#### PROCESS

- **Pre-analytic**
  - Ordering
- **Analytic**
  - Collection
  - Analysis
- **Post-Analytic**
  - Result Reporting/Display
  - Interpretation
  - Communication

#### Passive Display
- Prior results
- Cost
- Turnaround time
- Guidelines
- Web links

#### Interactive Alerting
- Alternatives
- Approval/second sign
- Duplicate checking
- Corollary orders
- High cost orders

#### Result Reporting
- Result flagging
- Result grouping
- Result structure
- Interoperability

#### Interpretive Guidance
- Automated coded comments
- Pathologist interpretations
- Guidance web links

Clinics in Lab Medicine, 39, 2019
EHR Lab Order Entry

- EHR order generation is complex: lab tests can be ordered via a facility list, departmental list, personal preference list, order set, therapy plan, or decision support rule.
- Team oriented medicine means many providers (resident, attending, medical assistant) may be involved in a single order.

Understanding EHR test ordering pathways is essential to control utilization.
Example: Rapid Responses to Volume Spikes

**RBC Folate**

- Large spike in RBC folate orders noted during routine monitoring
- With our Epic dashboard we were quickly able to localize the orders to a single enterprise anemia order set
- We first swapped out RBC folate for serum folate and later entirely removed RBC folate from the menu

---

I’m reaching out in regards to your ticket about switching the folate lab available in the anemia panel. I’ve chatted with our clinical content lead, and he wanted me to send along this screenshot for review:

> Folate levels in the liquid portion of blood (serum) can vary based on a person’s recent diet. Because red blood cells store 95% of circulating folate, a test to measure the folate level within RBCs may be used in addition to or instead of the serum test. Some health practitioners feel that the RBC folate test is a better indicator of long-term folate status and is more clinically relevant than serum folate, but there is not widespread agreement on this.

---

© 2019 College of American Pathologists. Materials used with permission of faculty.
Agile Approaches to Clinical Decision Support

- Laboratory is an integral part of the enterprise decision support team
- CDS team meets weekly to reset priorities for two week software sprints
- Has reduced the cycle time for decision support alerts to weeks instead of quarters
- Automation of monitoring with interactive, real time CDS dashboards
Example: CDS Alert to Discourage Immunofixation Ordering

Due to leveraging our existing EHR infrastructure the alert took < 5 hours of total effort including design, build, testing, monitoring and change management.

85% of the time providers accept the laboratory’s advice.

Annual savings of over $40,000 per year at a single hospital with this one alert.
Reusable dashboards and data models that are interactive and shorten the build cycle *(none built by the laboratory but many now heavily used by lab)*
New Approach: Sendout Adjusted Expenditure, Accounting for Diagnosis

- Instead of a pure focus on cost, worked with hospital economist who quantified the “expected cost” for a given outpatient visit based on EHR provider and patient characteristics
- Created P4P metric for neurology
- Using this approach across the board (labs, imaging, cardiology) for high expense testing at MGH

Cost in Excess of Expectation

© 2019 College of American Pathologists. Materials used with permission of faculty.
Information Processing in the Laboratory Testing Process

**PROCESS**

**Pre-analytic**
- Ordering
- Collection
- Automated Specimen Collection Process
- RFID/bar coding

**Analytic**
- Test Result Auto-verification
- Middleware
  - Interference checking
  - Rules-based auto-dilution
  - Automated add-ons
- Institutional Reflex Algorithms
- Enhanced Result Generation and Analytics

**Post-Analytic**
- Reporting
- Interpretable
- Info Buttons
  - Guidelines
  - Literature
  - Online resources
- Pathology Interpretative Services

**Enhanced Electronic Medical Record Systems**
- Actionable result reporting

© 2019 College of American Pathologists. Materials used with permission of faculty.
The Need for Informatics

- Millions of results per year
- Reported 24/7 with high levels of auto-verification
- Rate of data production exceeds capacity of clinicians, pathologists and technologists to generate information
- The human brain is not well equipped to process high dimensional data

Computational techniques needed
Recursive partitioning to “purify” data and generate intuitive decision trees

\[(x, Y) = (x_1, x_2, x_3, ..., x_k, Y)\]

\[I_C(f) = \sum_{i=1}^{m} f_i(1 - f_i) = \sum_{i=1}^{m} (f_i - f_i^2) = \sum_{i=1}^{m} f_i - \sum_{i=1}^{m} f_i^2 = 1 - \sum_{i=1}^{m} f_i^2\]
Pre-intervention, technologist judgment identified spurious results only 9% of time (9% sensitivity)

<table>
<thead>
<tr>
<th>Parameters Supplied in Building Tree</th>
<th>Data Set</th>
<th>Spurious Correctly Classified</th>
<th>Total Spurious</th>
<th>Real Correctly Classified</th>
<th>Total Real</th>
<th>Sensitivity (95% CI) %</th>
<th>Specificity (95% CI) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Na, K, Cl, Bicarb, Anion Gap, Glucose</td>
<td>Training</td>
<td>57</td>
<td>64</td>
<td>84</td>
<td>92</td>
<td>89 (79-95)</td>
<td>91 (84-96)</td>
</tr>
</tbody>
</table>

Implemented algorithm performed prospectively on real patients with 74% sensitivity and with 100% specificity
Machine Learning, Step by Step

Step 1: Frame Problem and Setup Data

<table>
<thead>
<tr>
<th>PT</th>
<th>Predictor A (e.g., BP)</th>
<th>Predictor B (e.g., Creat)</th>
<th>..</th>
<th>Clinical Outcome (e.g., progression to CKD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>137</td>
<td>1.36</td>
<td>..</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>168</td>
<td>1.19</td>
<td>..</td>
<td>Yes</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>..</td>
<td>...</td>
</tr>
</tbody>
</table>

Step 2: Clean Data

Step 3: Partition into training and testing

Step 4: Select type of model

Step 5: Train Model

Training Data Partition

Step 6: Test Model

Testing Data Partition

Performance Metrics

Prediction Algorithm

Step 7: Iterate

Step 8: Translate into a clinical algorithm

Steps 1-3, 8 are what you are already doing. Steps 4-7 may require a resource, but that resource can be YOU!
# Model Selection

## Selected Options

### Linear methods
- Ordinary least squares regression
- Logistic regression
- Perceptrons

### Decision trees
- Recursive portioning trees
- Ensemble methods (random forest)

### Artificial neural networks

### Support vector machines

### Neural Networks

## Considerations

- Data types
- Classification vs. regressions
- Complexity vs. data size
- Intuitiveness of output

- In practice often try multiple models
- Sometimes use an aggregate across various model types as final output

In prior example, we used recursive partition to produce interpretable decision trees
Machine Learning Based Multi-analyte Delta Checks Outperform Individual Analytes for Wrong Blood in Tube (WBIT) Detection

Positive Predictive Values at 80% Specificity

Prevalence

$$\begin{array}{c}
\text{Ca} & 0.01 \\
\text{Mg} & 0.07 \\
\text{BUN} & 0.05 \\
\text{Cr} & 0.02 \\
\text{Glu} & 0.03 \\
\text{Phos} & 0.01 \\
\text{AG} & 0.02 \\
\text{Cl} & 0.01 \\
\text{K} & 0.02 \\
\text{HCO}_3 & 0.14 \\
\text{Na} & 0.13 \\
\end{array}$$

© 2019 College of American Pathologists. Materials used with permission of faculty.

Am J Clin Pathol 2018;150:9
Machine Learning

- Clinical protocols developed using machine learning techniques have improved the laboratory’s identification and annotation of spurious results, anomaly detection, and WBIT errors.

- We have used similar techniques to develop EHR implementable rules for ordering alerts
  - e.g. Used machine learning to develop algorithm to suggest discontinuing peripheral blood flow cytometry orders when not indicated

- In addition to test result interpretation, machine learning can be used to predict test results

© 2019 College of American Pathologists. Materials used with permission of faculty.
Why Predict Lab Results?

1. **Eliminate redundant testing:** Tests that can be accurately predicted to be normal or abnormal may not be needed (*improve utilization*)

2. **Detect anomalies:** When the actual results are discrepant from predicted results, investigate and may report test results with a comment (*improve interpretation*)

3. **Avoid overlooked diagnoses:** Alert clinicians when tests are not ordered but are predicted to be abnormal (*avoid missed diagnosis*)
We used Ferritin in a proof-of-concept for test prediction

- Ferritin
  - A marker of iron stores
  - Used in the diagnosis of iron deficiency
  - Must be interpreted in the setting of other clinical and laboratory data
    - Decreased in iron deficiency
    - Increased in inflammation

“(Lab test) prediction is difficult, especially about the future”
Ferritin Methods Overview

Raw Data (3 months outpatient ferritin values)
- Transform ferritin values
- Divide at random into training and test partitions
- Mask ferritin results for test partition

Raw Data with ferritin masked for test patients
- Impute missing data
- Use 4 different imputation methods

“Completed” dataset
- Predict ferritin results
  - 4 regression methods
  - 1 classification method
  - Pair each with each imputation method

Performance Metrics
- Compare predicted ferritin results to measured results
- Compare predicted to “masked” results for test partition
- Review selected cases to determine clinical significance

Predicted ferritin results and classifications
Ferritin Classification Performance

Accuracy of Predicted Ferritin Results

>96% of the time we can accurately predict the ferritin without doing the test.


© 2019 College of American Pathologists. Materials used with permission of faculty.
Ferritin: Case Review

Conclusion: Predicted ferritin may more accurately reflect underlying iron status in some patients → potential application to clinical decision support

<table>
<thead>
<tr>
<th>Case</th>
<th>Ferritin</th>
<th>Predicted Ferritin</th>
<th>Impression</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>230</td>
<td>21</td>
<td>Iron deficiency, not clinically identified</td>
<td>Ferritin increased secondary to inflammation</td>
</tr>
<tr>
<td>2</td>
<td>197</td>
<td>19</td>
<td>Recovering iron deficiency</td>
<td>Receiving IV iron therapy</td>
</tr>
<tr>
<td>3</td>
<td>1768</td>
<td>9</td>
<td>Limited predictive data</td>
<td>Only two predictor tests available • Decision support will likely require a minimum number of predictor tests</td>
</tr>
<tr>
<td>4</td>
<td>197</td>
<td>19</td>
<td>Complex hematologic picture</td>
<td>Referral to hematology would have likely been useful had the testing been ordered by a non-specialist</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.
1. **Detecting anomalies:** Report test results with a comment when discrepant from predictions

2. **Avoiding overlooked diagnoses:** Alert clinicians when tests are not ordered but predicted to be abnormal

3. **Eliminating redundant testing:** Tests that can be accurately predicted may not be needed

---

This ferritin result is inconsistent with other testing. Do not exclude a diagnosis of iron deficiency on the basis of the ferritin alone.

Test results indicate the possibility of iron deficiency. Consider ordering ferritin if clinically indicated.
But what about the temporal component?

Real Data

Variables (Analytes)

Patients

Time points

Single Time Slice

VS.

Patients

Variables

Time points

<table>
<thead>
<tr>
<th>Pt</th>
<th>Age</th>
<th>Gen</th>
<th>Hct</th>
<th>Plt</th>
<th>...</th>
<th>K</th>
<th>Fer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>F</td>
<td>37.2</td>
<td>437</td>
<td></td>
<td>3.9</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>76</td>
<td>M</td>
<td>29.1</td>
<td>68</td>
<td></td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Test pt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>48</td>
<td>F</td>
<td>41.3</td>
<td>212</td>
<td></td>
<td>4.1</td>
<td>221</td>
</tr>
</tbody>
</table>
Many existing methods for analyzing time series are not well-adapted to laboratory data. Clinical laboratory data is irregularly and often sparsely sampled, of high dimensionality, and not sampled at random.
Develop the methods and models to accomplish the following:

**INPUTS**
For a given patient, time point and “analyte of interest”
- Prior values for the analyte of interest and other analytes
- Values for other analytes at the same time
- Desired confidence level ($\alpha$)

**MODEL**

**OUTPUT**
- Point estimate and confidence interval for the analyte of interest

**Example**
Mr. Smith’s prior test results

HCT Model

Mr. Smith’s Predicted HCT

© 2019 College of American Pathologists. Materials used with permission of faculty.
2 patients with Na, K, Cl measured on days 0-3

- Day 4 values = imputations with confidence intervals
- Patient 1 might not need this testing
- Patient 2 needs day 4 electrolyte testing since values may be abnormal
Traditional Reference Ranges are Inadequate for AKI Detection

Am J Clin Pathol. 2015, 143:42.

© 2019 College of American Pathologists. Materials used with permission of faculty.
Using 3D-MICE, Future Lab Tests Can Be Predicted With High Accuracy Using Trends and Ancillary Data

Narrow prediction width

High sensitivity and specificity

“Tomorrow's creatinine”


© 2019 College of American Pathologists. Materials used with permission of faculty.
Accurate Prediction Can Permit Earlier Intervention

- Instead of alerting providers that their patients are **now in early acute kidney injury (AKI)**, alert providers that in 24 hours their patients will likely satisfy criteria for AKI.
- Alert suppressed if calculation of prediction interval width is too wide.
- Requires real time data access for data analysis.

“Tomorrow's creatinine”

AKI Prevention Trial: Combining Analytics with Traditional Laboratory Evaluations

- New biomarkers for acute kidney injury (e.g. NGAL) may be accurate but are high cost and indications are unclear
- If we add to the EHR inpatient menu we can guarantee ourselves a $300,000 cost for testing without clear evidence of improved outcomes
- **Hypothesis**: One way of efficiently using new biomarkers may be to add them on as a reflex test when prediction criteria have been satisfied
- In the current study when the algorithm predicts a patient will likely meet renal failure criteria in 24 hours → Add NGAL to assess for kidney injury now

“Tomorrow's creatinine”
Informatics can provide the tools to address a wide variety of clinical and operational issues

- Pathologists should engage with and become educated in the capabilities of their EHR
- Most groups already have the infrastructure, pathologists need to step up to the EHR table
- Partnering with computational colleagues can provide mutual benefit but learning the basics yourself is important to be able to frame the questions you want to ask
A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD

© 2019 College of American Pathologists. Materials used with permission of faculty.
“WSI Meets Deep Learning: A Preview of the Future

Michael Feldman, MD, PhD
feldmanm@mail.med.upenn.edu
@feldmanm30
Professor University of Pennsylvania
Vice Chair Pathology & Laboratory Medicine
## Digital Workflow and ML/AI

<table>
<thead>
<tr>
<th>Microscope</th>
<th>Res/Fellows</th>
<th>Attending</th>
<th>Molecular IHC</th>
<th>Manual</th>
<th>Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue</td>
<td>Slide</td>
<td>Screening</td>
<td>Predictive</td>
<td>Quantitation</td>
<td>Dx</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rescreening</td>
<td>IHC</td>
<td>IHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Automated quant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Companies</td>
<td>Rescreening</td>
<td></td>
</tr>
<tr>
<td>HistoQC</td>
<td>Event finding:</td>
<td>IBRIS:</td>
<td>Automated quant</td>
<td>Rescreening</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Ca in Nodes</td>
<td>- Breast</td>
<td>- Companies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Ca in tissue</td>
<td>- Lung</td>
<td></td>
<td>EHR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mitoses</td>
<td>- Prostate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Grading</td>
<td>- ENT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Missed events:**
- Neg Bx - relook

---

© 2019 College of American Pathologists. Materials used with permission of faculty.
Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study)

The American Journal of Surgical Pathology: November 2017
MSK Whole slide imaging equivalency and efficiency study: experience at a large academic center
Mod Path Feb 2019, Sirintrapun et al

Table 3 List of specimens in each respective subspecialty

<table>
<thead>
<tr>
<th>Breast</th>
<th>Genitourinary</th>
<th>Gynecologic</th>
<th>Gastrointestinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>Prostate</td>
<td>119 Cervix</td>
<td>13 Stomach</td>
</tr>
<tr>
<td>Lymph node</td>
<td>28 Bladder</td>
<td>19 Uterus</td>
<td>10 Colon</td>
</tr>
<tr>
<td>Other</td>
<td>6 Kidney</td>
<td>8 Fallopian tube</td>
<td>9 Rectum</td>
</tr>
<tr>
<td></td>
<td>Ureter</td>
<td>1 Ovary</td>
<td>4 Gallbladder</td>
</tr>
<tr>
<td>Bone and soft tissue</td>
<td>Urethra</td>
<td>2 Vulva</td>
<td>2 Liver</td>
</tr>
<tr>
<td>Soft tissue</td>
<td>12 Testis</td>
<td>4</td>
<td>Pancreas</td>
</tr>
<tr>
<td>Bone</td>
<td>9 Lymph nodes</td>
<td>4 Dermatopathology</td>
<td>Esophagus</td>
</tr>
</tbody>
</table>

Table 5 Summary of interobserver concordance and turnaround times of glass and digital reporting

<table>
<thead>
<tr>
<th>Equivalence</th>
<th>Efficiency*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>99.3%</td>
</tr>
<tr>
<td>Grade</td>
<td>94.10%</td>
</tr>
<tr>
<td>Margin</td>
<td>100%</td>
</tr>
<tr>
<td>LV/LPI</td>
<td>83.80%</td>
</tr>
<tr>
<td>pT</td>
<td>97.30%</td>
</tr>
<tr>
<td>pN</td>
<td>97.10%</td>
</tr>
<tr>
<td></td>
<td>Pathologist H 5:02:30</td>
</tr>
</tbody>
</table>

*Median time difference was 19s longer per whole slide image and 47s longer per digital image.
Unmet Need

➢ Transition to digital pathology workflows
  – Digital Quality Control is paramount
  – Recut and rescan slides immediately before getting to a pathologist
  – Cost and efficiency savings

➢ Previously not insurmountable
  – Increasingly too time consuming to do manually
  – Non-reproducible

Slides taken from diagnostic cohort of TCGA-BRCA

We need better quality control of our slides!
HistoQC Properties

- Fast: n=1,143 in 466 minutes (24s/slide) using 6 cores (1.1TB)
- Easy “install” (git clone) with minimum dependencies:
  - python-openseslide, matplotlib, numpy, scipy, skimage, sklearn
- User interface is a single local html5 + JS file, no hosting
- No specialized hardware
- No internet connection required
- Designed to be modular and easily extendible
HistoQC...Your Pixels Matter

HistoQC: reproducible slide quality metrics with artifact localization

github.com/choosehappy/HistoQC

andrew.janowczyk@case.edu
andrewjanowczyk.com

HistoQCREpo.com
Gold in the hills…
Role of Tumor Morphology ER+ Breast Ca

• Modified Bloom-Richardson (mBR) grading (Elston and Ellis, 1991)
  – Tubule formation, nuclear pleomorphism, mitotic activity
• mBR identifies tumors as low, intermediate and high grade.
• Correlation between tumor Grade and outcome
• Visually determined, qualitative
• High inter- and intra-observer variability
  – Among 7 pathologists: $k = 0.50 – 0.59$ (Meyer et al., 2005)
  – Between pathology departments: $k = 0.51$ to $0.54$ (Boisen et al., 2000)
• Suboptimal treatment can result from incorrect grading (Dalton et al., 2000)
IbRiS: Comparing against Oncotype Dx RS

~450 feature data space
Hand crafted features
Inspired by Pathology expertise

<table>
<thead>
<tr>
<th>Good vs. poor</th>
<th>Good vs.</th>
<th>Intermed. vs. poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.91 ± 0.022</td>
<td>0.76 ± 0.051</td>
</tr>
<tr>
<td>PPV</td>
<td>0.94 ± 0.10</td>
<td>0.85 ± 0.11</td>
</tr>
</tbody>
</table>

PPV - Positive predictive values
Ibris and outcomes ECOG 2197

<table>
<thead>
<tr>
<th>Assay</th>
<th>% of patients with no recurrence after 10 years classified as low-risk</th>
<th>10-year recurrence rate in low-risk group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibris</td>
<td>57.5%</td>
<td>17.2%</td>
</tr>
<tr>
<td>Odx</td>
<td>56.3%</td>
<td>20.0%</td>
</tr>
<tr>
<td>Ibris + Odx</td>
<td>65.0%</td>
<td>16.0%</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.
Findings Cancer – Deep learning

Nature Sci Reports 2017 Apr 18;7:46450; Madabhushi et al
Breast cancer maps

False positives
a. DCIS
b. Sclerosing lesions

False negatives
a. Small areas of invasion

Table 1. Performance measures for the ConvNet classifier on the TCGA (pathological, N = 195) and NC (normal, N = 21) data cohorts. The measures included Dice, PPV, NPV, TPR, TNR, FPR, and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Dice</th>
<th>PPV</th>
<th>NPV</th>
<th>TPR</th>
<th>TNR</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA</td>
<td>0.7586 ± 0.2006</td>
<td>0.7162 ± 0.2204</td>
<td>0.9077 ± 0.0511</td>
<td>0.8691 ± 0.1582</td>
<td>0.9218 ± 0.0764</td>
<td>0.0782 ± 0.0764</td>
<td>0.1309 ± 0.1582</td>
</tr>
<tr>
<td>NC</td>
<td>N/A</td>
<td>N/A</td>
<td>1 ± 0</td>
<td>N/A</td>
<td>N/A</td>
<td>0.9664 ± 0.0110</td>
<td>0.0036 ± 0.0110</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of author.
Finding Lymph node mets ISBI 2016
CAMELYON 2016/17
11 Pathologist with time constraint or without time constraint compare to different machine learning algorithms (23 teams, 32 models)

<table>
<thead>
<tr>
<th>Data Set (N = 399 Slides and Images)</th>
<th>Hospital Providing the Slides and Images</th>
<th>Primary Tumor Histotype IDCA</th>
<th>Slides Containing Metastases, No.</th>
<th>No. of Lesions per Slide or Image, Median (Range)</th>
<th>Total Slides or Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (n = 270 images)</td>
<td>RUMC</td>
<td>IDC 54</td>
<td>None 100</td>
<td>2 (1-20)</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>UMCU</td>
<td>Non-IDC 16</td>
<td>Macro 60</td>
<td>3 (1-27)</td>
<td></td>
</tr>
<tr>
<td>Test (n=129 slides and images)</td>
<td>RUMC</td>
<td>IDC 23</td>
<td>Macro 50</td>
<td>2 (1-14)</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>UMCU</td>
<td>Non-IDC 6</td>
<td>Micro 30</td>
<td>3 (1-25)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CAMELYON16, Cancer Metastases in Lymph Nodes Challenge 2016; IDC, infiltrating ductal carcinoma; RUMC, Radboud University Medical Center; UMCU, University Medical Center Utrecht.

* All analyses in the training set were determined with whole-slide images.

* Primary tumor histotypes included IDC and other histotypes (non-IDC).
### Machine vs person Performance (ROC)

<table>
<thead>
<tr>
<th>Model</th>
<th>Performance (ROC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologist</td>
<td>0.966 (3.4% misses) WOTC</td>
</tr>
<tr>
<td>Harvard (Best algorithm)</td>
<td>0.925 (7.5% misses)</td>
</tr>
<tr>
<td>Combination Man + Machine</td>
<td>0.994 (0.6% misses)</td>
</tr>
<tr>
<td>Dual Neural net</td>
<td>0.994</td>
</tr>
</tbody>
</table>

**FROC** = sensitivity at various FP rates

@8FP is FN rate at 8FP per slide

---

**Table 1. Results on Camelyon16 dataset (95% confidence intervals, CI). Bold indicates results within the CI of the best model. “Small” models contain 300K parameters per Inception tower instead of 20M. -: not reported. *A pathologist achieved this sensitivity (with no FP) using 30 hours.**

<table>
<thead>
<tr>
<th>Input &amp; model size</th>
<th>Validation</th>
<th>Test</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50X</td>
<td>98.1</td>
<td>91.1</td>
<td>96.7</td>
</tr>
<tr>
<td>50X-pretrained</td>
<td>90.3</td>
<td>91.1</td>
<td>95.7</td>
</tr>
<tr>
<td>40x-small</td>
<td>93.3</td>
<td>92.4</td>
<td>97.7</td>
</tr>
<tr>
<td>ensemble-663</td>
<td></td>
<td>92.4</td>
<td>97.7</td>
</tr>
<tr>
<td>20X-small</td>
<td>94.7</td>
<td>91.1</td>
<td>95.6</td>
</tr>
<tr>
<td>10X-small</td>
<td>88.7</td>
<td>84.9</td>
<td>96.5</td>
</tr>
<tr>
<td>40X+20X-small</td>
<td>94.9</td>
<td>92.0</td>
<td>97.0</td>
</tr>
<tr>
<td>40X+10X-small</td>
<td>93.8</td>
<td>87.0</td>
<td>98.6</td>
</tr>
<tr>
<td>Pathologist [1]</td>
<td></td>
<td>91.9</td>
<td>99.9</td>
</tr>
<tr>
<td>CAMELYON16 winner [2,3]</td>
<td></td>
<td>80.7</td>
<td>82.7</td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.
Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer

AJSP 42(12): 2018 Stumpe et al
Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Prostate Pathology

12,000 WSI

Fuchs et al ArXiv May 2018
Maccabi Healthcare Services is a large healthcare provider with a centralized pathology institute - 120,000 histology accessions per year
- ~700 prostate core needle biopsies (PCNBs)
- Roughly 40% of the PCNBs are diagnosed with cancer.

IBEX Medical Analytics, whole slide images of PCNBs, including cancerous glands (of Gleason patterns 3, 4 and 5), high-grade PIN and inflammation. The algorithm utilizes state-of-the-art Deep learning CNN, trained on many thousands of image samples, taken from hundreds of PCNBs from multiple institutes, and manually annotated by senior pathologists.

Small study shown at ECDP 2018 in Helsinki – 100 retrospective cases that had been diagnosed as benign, and found two three errors
- In two cases, the algorithm identified small foci of Gleason 3. Placed into watchful waiting groups. Two years later, both patients were diagnosed with higher grade cancer and underwent radical prostatectomy.
- Third case was a larger focus of pseudo-hyperplastic CAP, resection showed a CAP(4+3) confined to prostate.

System now used to rescreen all negative core prostate biopsies
- New workflow, AI has 30-40% false positive rate – pathologist then reviews specific cores identified by hotspots to decide if any lesion needs further workup or staining
Problems with AI

Brittle

- Most published papers are small data sets that don’t always validate well
  - Deep Mitoses – training vs validation very different

Table 4
Performance comparison of DeepDet with other competing approaches on 2012 MITOSIS test set.

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepDet</td>
<td>0.854</td>
<td>0.812</td>
<td>0.832</td>
</tr>
<tr>
<td>RRF (Paul et al., 2015)</td>
<td>0.835</td>
<td>0.811</td>
<td>0.823</td>
</tr>
<tr>
<td>CasNN (Chen et al., 2016a)</td>
<td>0.804</td>
<td>0.772</td>
<td>0.788</td>
</tr>
<tr>
<td>HC-CNN (Wang et al., 2014)</td>
<td>0.84</td>
<td>0.65</td>
<td>0.735</td>
</tr>
<tr>
<td>IDSSA (Greşan et al., 2013)</td>
<td>0.886</td>
<td>0.70</td>
<td>0.782</td>
</tr>
<tr>
<td>IFAL (Ishad et al., 2013)</td>
<td>0.698</td>
<td>0.74</td>
<td>0.718</td>
</tr>
<tr>
<td>SUTECH (Tashk et al., 2013)</td>
<td>0.70</td>
<td>0.72</td>
<td>0.709</td>
</tr>
<tr>
<td>NEC (Malon et al., 2013)</td>
<td>0.75</td>
<td>0.59</td>
<td>0.659</td>
</tr>
</tbody>
</table>

Table 5
Performance results of our methods on 2014 MITOSIS validation set.

<table>
<thead>
<tr>
<th>Method</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepDet (fixed)</td>
<td>0.489</td>
</tr>
<tr>
<td>DeepDet+Seg</td>
<td>0.505</td>
</tr>
<tr>
<td>DeepDet+Seg+Ver(c)</td>
<td>0.559</td>
</tr>
<tr>
<td>DeepDet+Seg+Ver(f)</td>
<td><strong>0.572</strong></td>
</tr>
</tbody>
</table>

© 2019 College of American Pathologists. Materials used with permission of faculty.

Medical Image Analysis 45 (2018) 121–133

Fuchs et al ArXiv May 2018
Path Rads integration

See the bigger picture

Screening  Ultrasound  Biopsy  Surgery  Specimen  Sentinel  Review  Treatment  Follow up

High volume
- Workload balancing
- Access to subspecialists

Common Workflow
- Synoptics
- Data from Image
- Compute on Image
- FTE
- Server/Storage
- EMR/LIS interface
- AI/ML

Low volume
- Consultations to anyone
- Second opinion

Central PACS

Image Exchange Portal

© 2019 College of American Pathologists. Materials used with permission of faculty.
Industry timeline digital pathology

- **Gen1**: 1998, 60 min scan, $$$
- **Gen4-5**: 2017, FDA
- **Gen6**: 2018, <60 s scan, Z axis coming, $$$$

**V1.0**: Computational Photonics
- 5 min
- Z axis
- $

**V1.5**: Slidefree imaging
- 2-3 min
- experimental

**V2.0**
New imaging modalities coming
Light-sheet Fluorescence Microscopy

- Separate excitation and emission paths
- Sheet of excitation created by deep depth of focus
- Increased speed with imaging a plane rather than single point
- Able to reconstruct optical sections

Liu et al, Nature Biomed 2017
Invasive Ductal Carcinoma
Digital Pathology is not just images

- Treanor et al 2014 time and motion studies
  - 1/4-1/2 of time is looking through medical record for data
  - How do we make data gathering more precise, more focused and faster?
  - Center healthcare Innovation EHR extensions
    - Yevgeniy Gitelman, Katherine Choi
    - Oncology, Pathology, Radiology
Focused Snapshots: Breast Oncology

Reduce Hunting for Information

Baseline: Time to prep chart before visit = 30 min / new patient referral

Notes serve many purposes

For future self:
Avoid the pain of recreating the picture of the patient from scratch

For others:
Show what’s relevant at time of note being written

Gathering relevant data to form a picture of the patient is effortful

Too much screen switching between data types
Drilldown is inefficient: reports and links are buried
Difficult to jump into timeline where most relevant
Gaps: Others’ notes focus on diff parts of the story
Gaps: Feeling there is a clue in the chart you didn’t see
Lots of reformatting
Unclear what data other providers need
Exploring EHR Extensions

Notes, Imaging, and Surgical Pathology on one timeline chronology

Focus on the key 30%

Smart previews show “Impression” for quick scanning

Focus on the key 5%

Full reports for drilldown
Eliminate non-clinical 50%

© 2019 College of American Pathologists. Materials used with permission of faculty.
Conclusions

- FDA clearance is only a beginning
- Machine learning will accelerate
  - Targeted review
    - Rare event detection
    - Tumor finding
    - Feature classification
    - Grading
    - Screening/Rescreening
  - Outcome prediction
  - qIHC, qMultiplex
- Large well curated and annotated datasets are platinum
- Data Science is our future
  - AI and ML are key to unlock our data
  - Path-Rads integration is our future
  - New technologies coming
- Business – new models of practice
Case Western Lab Director: Anant Madabhushi, PhD

Postdocs: James Monaco, PhD  
Gaoyu Xiao, PhD  
Jun Xu, PhD  
Andrew Janowczyk, PhD

Graduate Students: Jonathan Chappelow  
Scott Doyle  
Satish Viswanath  
Pallavi Tiwari  
George Lee  
Shannon Agner  
Ajay Basavanhally  
Rob Toth  
Andrew Janowczyk

Undergraduate Students: Jay Naik  
Hussain Fatakdwala  
Amod Jog

Penn Clinical Collaborators: Mitch Schnall, MD, PhD  
David Roth, MD, PhD  
John E Tomaszewski, MD  
William Lee, MD  
Natalie Shih, MD

Clinical Collaborators: Shridar Ganesan, MD, PhD

Penn Center Clinical Innovation: Roy Rosin  
Yevgeniy Gitelman, MD  
Katherine Choi

© 2019 College of American Pathologists. Materials used with permission of faculty.
Introduction To Machine Learning Using Examples From Anatomic Pathology

September 24, 2019

Andrew Janowczyk, PhD
Assistant Research Professor

© 2019 College of American Pathologists. Materials used with permission of faculty.
Outline

➢ What are images?
➢ What can be done with them?
➢ Feature extraction
➢ Intuition behind Classifiers
  – Real world examples
➢ Important considerations
  – Types of annotations
  – Batch effects
  – Quality Control

➢ If anything is unclear, let me know!
Digital pathology images are pixels
Images are 3D matrices of Pixels
Width x Height x [Red Green Blue]
Pixels can have operations done on them

- **Processing and Filtering:**
  - Taking a pixel
  - Perform a function
  - New Value

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>.3</td>
<td>.5</td>
</tr>
<tr>
<td>.3</td>
<td>1</td>
<td>.2</td>
</tr>
<tr>
<td>.6</td>
<td>.3</td>
<td>.9</td>
</tr>
</tbody>
</table>

> .5

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Can convert from RGB to Grayscale

- Pixel by pixel - Linear equation
- Gray = 0.2989 * R + 0.5870 * G + 0.1140 * B
Can apply a threshold

- Images range from [0 = black, 1 = white]
- Values < .5
Pixels can have operations done on them

- **Processing and Filtering:**
  - Taking a pixel
  - Look at its neighbors
  - Perform an operation

### Processing Examples

<table>
<thead>
<tr>
<th>2</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

**Sum**

40

<table>
<thead>
<tr>
<th>?</th>
<th>-1</th>
<th>-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>6</td>
<td>-1</td>
</tr>
<tr>
<td>?</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Can reduce noise by smoothing
➢ Each pixel is replaced by the mean value around it
Can look at values as a collection

- Precisely quantify familiar image properties

Average Brightness: 140.7963
Can look at values as a collection
➢ Precisely quantify familiar image properties

Average Brightness: 114.5927
Filtering and Processing using operations

➢ For example, can identify edges by subtracting adjacent pixels
➢ $0 - 0 = 0$, $1 - 1 = 0$, $1 - 0 = 1$, $0 - 1 = 1$
Example edge detection in DP space
These outputs lead to features

- Domain relevant
- Used in grading schemes
- Properties
  - Size
  - Shape
  - Texture
- More precise quantification as features
  - Eccentricity (how circular)
  - Length major/minor axis
  - Orientation
  - Staining intensity
  - Smoothness
  - Entropy
Texture Features

- Stroma region is “smooth”
- Measure of homogeneity
- Neighboring pixels are similar
- Small gradient between them

- Other regions more “rough”
- Measure of heterogeneity
- High amounts of entropy
- Larger and unpredictable gradients
Graph Features

- Can think of it in terms of connectivity
- At the object level:
  - How many neighbors do I have in a defined radius?
  - Average length away?
- Measurement for infiltration
  - How far am I away from a boundary?
  - How far am I away from a cancer cell?
Remarks on features

➢ Try to avoid throwing the “kitchen sink” at a problem
➢ Start with a “reasonable” subset
➢ Based on:
  – Domain expertise
  – Grading schemes
Approaches using features

- Active contours
- Keep expanding boundary of an initial box while:
  - Inside is homogenous
  - No edge detected
What can we do with features?

➢ Measure difference between classes
➢ Quantify differences:
  – Benign vs Malignant
  – Subtypes
  – Outcome
  – Therapy Response
Feature extraction
➢ Goal of the model
  – Fit training data well
  – Generalize to Testing data
➢ If identified something biologically relevant and “true”
  – should be consistent
➢ Can we improve the model by adding dimensions?
This problem only gets worse

- The more dimensions, the larger the solution space is
- A lot of noise as well
  - Measurements
  - Labels
- Optimization is hard and time consuming
- Is there anything we can do to help?
Circularity

Small Area Large

Hopefully doesn’t occur “in nature”
Difference Machine vs Deep Learning?

➢ Machine Learning:
  – Explicitly provide feature measurements: e.g., Area, Circularity

➢ Deep Learning:
  – Self-discover the features

➢ Both approaches require good quality examples
What kind of examples are best?

- Near to decision boundary!
- Information rich
- Cancer vs non-cancer

Don’t need 10000s of these!!!
Examine use cases

- Segmentation
- Feature Extraction
- Learn Model
Cell Orientation Entropy (COrE) Features Stratify More and Less Aggressive Prostate Cancer on Tissue Microarrays

Aggressive cancer (left) shows more disorder in orientation of the nuclei compared to less aggressive cancer (right).

Cell Cluster Graph for Prediction of Biochemical Recurrence in Prostate Cancer Patients from Tissue Microarrays

- Novel Cell Cluster graph (CCG) that can quantify tumor morphology
- Extracted features from CCG can predict Biochemical recurrence in Prostate Cancer in 80 patients.

<table>
<thead>
<tr>
<th>Voronoi</th>
<th>Delaunay</th>
<th>CCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.1 ± 1.8%</td>
<td>60.7 ± 0.9%</td>
<td>83.1 ± 1.2%</td>
</tr>
</tbody>
</table>

Table 2. Comparison of CCG against other graph based methods in predicting biochemical failure.

Computerized nuclear features predict recurrence in lung cancer

Spatial arrangement of tumor infiltrating lymphocytes (TILs) predict response to Nivolumab in non-small cell lung cancer (NSCLC)

**Hypothesis:** Spatial arrangement of TILs and local density variance are highly correlated to the patient response.

**Data sets:**
Two independent data (whole slide image) acquired from UPenn (32) and CCF (24)

**TIL detection and image feature extraction**

TILs (green) & Non-TIL(Yellow)

**Top 5 most significant features obtained by feature selection**
1. Median of TILs formed areas
2. Ratio of Cancer cells to TILs cells
3. Cancer cell averaged Density
4. Density of TILs
5. Median of Cancer cell formed areas

A QDA classifier was trained using a Training set (n=32) and a independently validation set from a different institution (n=24).

Outline

➢ What are images?
➢ What can be done with them?
➢ Feature extraction
➢ Intuition behind Classifiers
➢ Important considerations
   – Types of annotations
   – Batch effects
   – Quality Control
Types of annotations and tasks

➢ Detection
➢ Bounding box
➢ Segmentation
Types of annotations and tasks

➢ Detection – Where is it?
➢ Typically place dot in center
➢ Pros:
  – Fast and easy
  – Easy to score
➢ Cons:
  – No size information
  – No shape information
➢ Use cases:
  – Mitosis detection
  – Lymphocyte detection
Types of annotations and tasks

➢ Bounding box – Where and about how big is it?
➢ Smallest box which will surround object

➢ Pros:
  – Give size information
  – Faster than segmentation

➢ Cons:
  – Still no shape information

➢ Use cases:
  – ROI identification
Types of annotations and tasks

➢ Segmentation – What are its borders?
➢ Precisely circle object at a pixel level
➢ Pros:
  – All types of analysis are possible
  – Morphological analysis
➢ Cons:
  – Very time consuming
  – Lots of noise: human error + ambiguity
➢ Use cases:
  – Nuclei segmentation
60 minutes to gather annotations, how?

➢ Many images sparsely annotated
  – High image number, low density
➢ A few images fully annotated
  – Low image number, high density
➢ Multiple images with selected ROIs fully annotated
  – Modest image, modest density
➢ Last approach is the best!
  – More patient diversity
  – More region diversity
  – Likely to be “different” and informative
How to select the ROIs?

➢ Exploit domain knowledge
  – I know nuclei + lymphocytes appear similarly, try to find ROIs which have both present to challenge the classifier

➢ False positive sampling
  – Train a model
  – Use it on training data
  – See where errors occur
  – Hyper-sample those types of regions

➢ Ultimately, the classifier can tell you where its struggling by displaying poor performance!

➢ Target those types
  – Similar to teaching a student

© 2019 College of American Pathologists. Materials used with permission of faculty.
Outline

➢ What are images?
➢ What can be done with them?
➢ Feature extraction
➢ Intuition behind Classifiers
➢ Important considerations
  – Types of annotations
  – Batch effects
  – Quality Control
Batch Effects

➢ Confounding of non-biological signal with biological signal

Group 1

Group 2

Where should these go?
Batch Effects 2

- You’ve done this grouping by color and appearance
- Not taken into account any biological information e.g., disease presentation
- Examine in practice
- Pathologist marked slide with dot
- Identify/scan (rare) samples
- Add in “undotted” samples
- Classifier learned to focus on dot
- Great performance on test set!
- Very poor performance on external set!
How would Batch Effects present in DP

- Pre-scanning: stain intensity, thickness
- Scanning: brightness, compression, microns per pixel
Potentially Likely (Worst) Situation

- TMA created containing only high-risk patients
- TMA created containing only low-risk patients
- *Any* artifact will perfectly separate groups
Unmet Need For Quality Control

➢ Transition to digital pathology workflows
  – Digital Quality Control is paramount
  – Recut and rescan slides immediately before getting to a pathologist
  – Cost and efficiency savings

➢ Previously not insurmountable
  – Increasingly too time consuming to do manually
  – Non-reproducible

Slides taken from diagnostic cohort of TCGA-BRCA

We need better quality control of our slides!
What is HistoQC?

- Open source reproducible slide quality metrics with artifact localization
- Python backend
  - identify artifacts and produce binary masks of “good” tissue
  - compute actionable quality scores and metrics
- HTML5 front end for visualizing and investigating results
- Received innovation award at European Congress on Digital Pathology 2018
End Users of HistoQC

➢ Pathology Departments
  – Real-time rolling average of metrics
  – Identify issues early

➢ Repositories + Computational pathologists
  – Identify and avoid artifacts and outliers for better datasets
    • Stain variances
    • Micron per pixel (MPP) heterogeneity
    • Batch effect presence
  – Explicitly define acceptable tolerance
Example Feature: Template Matching

<table>
<thead>
<tr>
<th></th>
<th>Template 1</th>
<th>Template 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Slide</td>
<td>0.00009</td>
<td>0.00212</td>
</tr>
<tr>
<td>Right Slide</td>
<td>0.00485</td>
<td>0.00092</td>
</tr>
</tbody>
</table>

Is slide within tolerances? Will my algorithm work?
• Precisely specify ranges
Quality Control Slide Repository

- Created website to host the “greatest hits”
- Slide and associated metadata (e.g., artifact type)
- Useful as a didactic tool for new pathologists
- Benchmark algorithms
  - Detecting artifacts
  - Measure algorithm robustness to artifacts
- Currently available: [http://www.histoqcrepo.com](http://www.histoqcrepo.com)
Final Call To Action - Download HistoQC

- Try on your data
- Submit pull requests for new modules

_HistoQC: reproducible slide quality metrics with artifact localization_

github.com/choosehappy/HistoQC

- Upload/download artifact containing slides to/from repository:
  HistoQCRepo.com

Thank you!

andrew.janowczyk@case.edu
andrewjanowczyk.com
Visualizing Individual Results
What to do with outliers?

Make a decision

1. Remove entirely from dataset if image is really bad
2. If rest of the image is okay, make sure to avoid that bad region
3. Can (and should) use HistoQC output mask to select regions to sample from!