Introduction to Computational Pathology

Fayyaz Minhas
Department of Computer Science
TIA Centre
PathLAKE Consortium

@fayyazhere
Theme of this talk

• **Workflow**
  • Digital vs. Computational Pathology

• **State of Cpath**
  • 5-star system for measuring the impact of Cpath
  • Including a taxonomy of computational problems in CPath

• **Challenges and Next Steps**
Computational Pathology

AI to help understand the cause, nature, origin and patterns of diseases for clinical decision making

How much fat?
45.9%
Why is applying ML to pathology a good idea?

- It is a very visual discipline
- Subjectivity
- Quantification
- Digitization as an enabling technology
  - Flexible working
  - Good concordance with slide based clinical decision making
- Pathologist Recruitment and aging workforce
- Identification of discriminatory features
- Learning from “Data at Scale”
- Information Integration
- Clinical Impact
- Commercial Impact
Workflow

Tissue Acquisition → Slide Preparation → Scanning → Quality Control → Pre-processing → Computational Algorithms

Glioblastoma (GBM) Samples → Stained Tissue Slide → Image Analysis

Re Stained Tissue Slide
Prediction Problems in CPath

Given training data, Can we predict:

Grade?
   *Regression*
Tumour Regions?
   *Segmentation*
Quantifying cells
   *Quantification*
Receptor Status?
   *Classification*
Mutation Status?
   *Classification*
Survival?
Drug Response?
Similar images?
   *Retrieval*
Biomarkers?
   *Pattern Discovery*
Measuring Cpath Advancement

Scientific Achievements

Technological Advancements

Clinical Deployment

Measuring Cpath Advancement

Commercial Development

Academic & Training Programmes

Regulatory & Ethical Acceptance
Measuring Cpath Advancement

Scientific Achievements

Disruptive advances

Technological Advancements

“Sensor” Technologies

Data Management

AI Technologies

Training pathologists & Computer Scientists for CPath

Changing clinical workflows

Collaborative & Online Working

Market/Investment size

Sustainability

Market growth

Commercial Development

Clinical Deployment

Measuring Cpath Advancement

Academic & Training Programmes

Regulatory & Ethical Acceptance

Patient Impact

Regulatory agencies

• Other Stakeholders
  • Pharma
  • AI Industry
  • Computing Industry
  • Academia
  • Personalized Medicine
Technological Advancements in CPath

- Current Focus
  - Image analysis with Machine Learning

“Sensor” Technologies
Data Management
AI/ML Technologies
Cpath “Sensor” Technologies

• Whole Slide Image Scanners
  • Cost, quality and throughput
• Upcoming Technologies
  • Multiplexed Imaging
  • *In Situ* Genomics and Transcriptomics
  • Multispectral Imaging
  • Liquid Biopsies
  • 3D Imaging of WSIs
  • Virtual biopsies and autopsies: Radomics
  • Non-destructive, *In Vivo* and RT pathology
  • Microscopic imaging in the body
    • Multi-photon microendoscopy

1 to 5 min for a small biopsy to 5–20 min for a surgical specimen and 3–5 min for a liquid-based cytology smear

“The cost of procurement, implementation, and operational costs of WSI may be prohibitive, especially for small pathology laboratories due to huge initial cost of the scanners (US $100,000 to US $1,500,000 per piece) and additional hidden costs of training of staff and pathologists, technical support, digital slide storage systems, and regulatory or licensing costs”

“WSI images would be only as good as the original glass slide”

“Sensor” technology challenges

• Reduction in scanning costs
  • Does it match the decrease in sequencing or transcriptomic analysis?

• Integration of upcoming imaging technologies in clinical and scientific discovery

• Open-source hardware?
  • Smartphone systems?

• Integration into LIMS

CPath Data Management

- Development of large-scale openly accessible well-annotated data repositories
  - TCGA, CPTAC, PathLAKE, BIG-PICTURE
- De-identification
- Data Transfer
- Datalake design
  - Cloud based analytics or On-Prem solutions?
- Interoperability
  - Image formats: svs, DICOM, ...
  - Annotations and pathology reports
- Collaboration and retrieval permissions
Technology Advances in ML for CPath

• Advances
  • Largely tracking advances in Computer Vision and Machine Learning

• Developed Algorithms for:
  • Pre-processing
  • Quality Control
  • Classification
  • Regression
  • Survival Prediction
  • Segmentation
  • Quantification/Counting
  • Object Detection
  • Semantic Segmentation
  • Contrastive Learning
  • Self-supervised learning
  • Active Learning
  • Registration, …
  • Generative Modelling
  • Virtual staining
Stain Normalization

The problem of ‘normalizing’ the stain colour distribution according to some ‘reference’ image
Stain Normalization
Stain Normalization
Classification

(a) tumour epithelium, (b) simple stroma, (c) complex stroma (stroma that contains single tumour cells and/or single immune cells), (d) immune cell conglomerates, (e) debris and mucus, (f) mucosal glands, (g) adipose tissue, (h) background.

(Kather et al. 2016)
Regression

- Cellular Composition Prediction: Counting how many different types of cells in a given patch

<table>
<thead>
<tr>
<th>Cell type/Method</th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Image 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplastic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammatory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epithelial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ground Truth</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>HoVer-Net</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>ALBRT</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>23</td>
</tr>
</tbody>
</table>


Object Detection and Classification


Nuclear Segmentation

One Model is all you need!

AI-Assisted Annotations

**NuClick** used for collecting the segmentation masks from point clicks

Koohbanani *et al.*, *Medical Image Analysis* (Oct 2020) & Gamper *et al.*, Arxiv (Apr 2020)
NuClick+: AI-Assisted Annotation of Glands

Koohbanani et al., Medical Image Analysis (Oct 2020)
Grading problems

Survival Prediction

- **Given**
  - A set of patients with their covariates and censored survival/recurrence data

- **Predict**
  - Survival curves and/or identify patterns associated with survival

Image Registration

Clustering: Unsupervised Learning
WSIs are graphs
WSIs are graphs.
WSIs are graphs
WSIs are graphs
Discriminative vs Generative Modelling

https://doi.org/10.1038/s43018-020-0085-8.
Virtual Staining

Advanced Machine Learning in CPath

• Contrastive Representation Learning
• Active Learning
• Self-Supervised Learning
• Graph Neural Networks
• Community Detection
• Attention-based modelling
• Super-resolution imaging
• ...
ML Technology in CPath

• Issues
  • Pre-processing and data quality management
  • Aggregation and information integration
  • Annotation Efficiency
  • Robustness
  • Explainability
  • Confidence prediction and “Rejection”
  • “Auto-ML” in CPath
  • Generalization and reproducibility
ML Problem(s) with WSIs: Multiresolution
ML Problem(s) with WSIs: Size
ML Problem(s) with WSIs: Spatial Context Matters
ML Problem(s) with WSIs: Training Labels

- Case or WSI-level labels
- Detailed annotations (Regions or Cells)

Ease of constructing ML models & Confidence

Annotation Volume

Case • WSI • Region • Cell • Subcellular

Annotation Detail

Case-level
- Normal
- Abnormal

Region-level
- Tumour
- Tumour associated strom
- DCIS
- Adenocarcinoma

Cell-level
- Tumour cells
- Stromal cells

Subcellular
- Immune cells
- Mitotic cells

Biopsy report

Pathology report

Patient name: [Name]
Date of birth: [Date]
Record number: [Number]

Specimen number: [Number]
Date obtained: [Date]
Date processed: [Date]

History: 62-year-old female with a palpable breast mass and suspicious calcification on mammogram.

Body site and procedure: Left breast at 9:00, 5 cm from nipple (ultrasound-guided core needle biopsy).

Description: Specimen is received in formalin, labeled with patient's name, and consists of 2 cores of tissue measuring 0.4 cm in width.

Diagnosis: Ductal carcinoma, Nottingham histologic grade 2. The carcinoma is positive for ER (80%) and is negative for HER-2 (IHC score 0).
Aggregation Problem

• Aggregation: Moving from patches to slide level predictions
  • Weak supervision approaches such as Averaging, Multiple Instance Learning, Attention Mechanisms LSTMs…
Weakly Supervised Learning Example

• Given:
  • Patches from Tumor Slides
  • Patches from Non-tumor slides

• Tumor Slides will have non-tumor patches

• Predict
  • Given a patch – is it a tumor patch?
  • Given a slide – does it have a tumor patch? (Is tumorous?)
An alternate modelling strategy: Graphs

• What are graphs?
  • Sets of nodes
    • Each node has a set of features
  • Links or Edges
    • Each edge can have a set of descriptors
WSIs are graphs

Generalization and Reproducibility

- Sources of variation
  - Inherent differences in populations
  - Difference in disease pathology
  - Tissue Processing and Handling
  - Scanner Changes
  - Pre-processing and QC
  - Changes in population
    - Due to treatment
    - Aging

- Possible Solution(s)
  - Publishing code and data
  - Preventing over-engineering
  - Large-scale validation
    - PathLAKE
  - Competitions
    - Mitosis detection
    - MSI prediction
  - Friendly “Adversarial” attacks
  - Peer-review process
  - Making solutions available to pathologists for trials (webservers)
If you can make a CPath Model, We can break it!

### Table: Platform Compatibility and Language Support

<table>
<thead>
<tr>
<th>Platform Compatibility</th>
<th>TIA Toolbox</th>
<th>HistoCarto Graphy(^\text{11})</th>
<th>HEAL(^\text{10})</th>
<th>QuPath(^7)</th>
<th>PathML(^\text{13})</th>
<th>CLAM(^\text{12})</th>
<th>Multi_Scale_Tools(^\text{14})</th>
<th>stainlib(†)</th>
<th>IBM CODAI deep-histopath(†)</th>
<th>HistoQC(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Python</td>
<td>Python</td>
<td>Python</td>
<td>Java ((+)Groovy)</td>
<td>Java ((+)Python)</td>
<td>Python</td>
<td>Python</td>
<td>Python</td>
<td>Python</td>
<td>Python</td>
</tr>
<tr>
<td>Modular Usage</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Unit Test Coverage</td>
<td>&gt;99%</td>
<td>88%</td>
<td>0%</td>
<td>87%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>?</td>
</tr>
<tr>
<td>Read Visual Fields (PNG, JPEG)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Read WSI TIFF (SVS, SCN, MRXS)</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Read Generic Tiled TIFF</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Read JP2 (Omnyx)</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Physical Unit Based Reads</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Tissue Masking / Detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patch Extraction</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stain Separation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stain Normalization</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Stain Augmentation</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patch Classification</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Tissue Semantic Segmentation</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Nucleus Detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Nucleus Segmentation</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Nucleus Classification</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Deep Feature Extraction</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Graph Learning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Whole Slide Classification</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

---

Future Technological Advances in CPath

• Overcoming previously stated challenges
• Incorporating Natural Language Processing
• Focus on information fusion

• Are pathologist labels noisy? (Invalidate Turing tests)
Scientific Achievements of CPath

• 5 Pathologist Nobel Prize Winners [1]
  • Johannes Fibiger received the Nobel Prize in Physiology or Medicine in 1926 for the experiments in which he produced gastric carcinoma in rats by feeding them Spiroptera-infected cockroaches.
  • George Whipple was awarded the 1934 Nobel Prize in Physiology or Medicine for his discovery that a diet rich in liver cured pernicious anemia.
  • Renato Dulbecco was awarded the 1975 Nobel Prize for discovering by molecular techniques that the genetic material of viruses was incorporated into the genetic material of the transformed cells and represented the first phase in carcinogenesis.
  • Baruj Benacerraf received the 1980 Nobel Prize in Physiology or Medicine for his discovery of immune response genes.
  • John Warren received the Nobel Prize in Physiology or Medicine in 2005 in recognition of this discovery of the role of H. Pylori in ulcers.

[1] http://www.annclinlabsci.org/content/39/2/196.long

“Rutherford used to divide science into physics and stamp collecting.” [2]
Scientific Discoveries of CPath

• What has CPath help discover?
  • Current focus
    • Identification of associations between various clinical parameters and Cpath Imaging data (visual and/or subvisual)

• It is possible to predict (from WSIs)
  • Mutations
  • Gene Expression Patterns
  • Source of cancers of unknown primaries
  • Microsatellite Instability
  • Receptor Status in BrCa
  • Chemo-response/drug response in OvCa
  • Recurrence
Some Literature


Mutation Prediction


Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer

Gene Expression Patterns

Receptor Status in BrCa

Cancers of Unknown Primaries

Chemotherapy Response Prediction


Future Scientific Discoveries of CPath?

• Previously undiscovered but clinically relevant patterns in data by integrating multimodal and multisource data
• Raising interesting biological questions
• Identification of novel causes of treatment effectiveness or resistance
• Leading to discovery of novel therapies and drugs
• Understanding evolutionary mechanisms at play that result in mutations
• Using Cpath as a tool by Pathologists
  • Current ML technologies are good at learning patterns
  • However, they do not have clinical and biological understanding of disease phenomenon
• Rare cancers

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6852275/
CCR Deployment

• How many commercial Cpath solutions have been clinically deployed (and are being used) after regulatory approval?
• Measuring Improvement in Clinical workflows and efficiency
• Closing the gap between volume of scientific publications and clinical deployment

A 510(K) is a premarket submission made to FDA to demonstrate that the device to be marketed is as safe and effective, that is, substantially equivalent, to a legally marketed device (section 513(j)(1)(A) FD&C Act) that is not subject to premarket approval.
Commercial Development

• Interesting questions: What business models work in Cpath?
  • Data to compute
  • Compute to data
  • Integrated frameworks vs. distributed tools
  • Public data lakes vs. proprietary data banks

• Estimating market size

<table>
<thead>
<tr>
<th>Market size value in 2021</th>
<th>USD 382.3 million</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue forecast in 2028</td>
<td>USD 593.9 million</td>
</tr>
</tbody>
</table>

Regulatory and Ethical Approval

- More data vs privacy
- “Black-box”
  - Bias?
- Data Repurposing?
- Competing interests?
  - Using data to sell a product?
- Rapidly changing regulatory landscape
  - European regulation (in vitro diagnostic medical device regulation) 2017/746 after May 26, 2024
Education and Training Programs

- How has Cpath improved training of pathologists?
- Improving recruitment into pathology as a discipline

- Lectures + Hands-on Workshops
  - Mon – Wed

- Hackathon
  - 5pm Wed – 4pm Fri

- Short Talks + Hands-on Workshops

- Panel Discussion

- Industry Exhibitions
Towards Reinventing Pathology with AI