Realizing Artificial Intelligence in Pathology

Streamlining Machine Learning Model Development and Clinical Deployment

Markus D. Herrmann, MD, PhD

June 10, 2021
Markus D. Herrmann, MD, PhD

- Interdisciplinary physician-scientist at the intersection of medicine, biology, and computer science with a specialization in biomedical imaging, machine learning, and clinical informatics.
- Studied medicine at Ulm University in Germany, where he also received his first PhD.
- Joined the faculty of MGH Pathology to lead the Computational Pathology division in 2019.
- Research focuses on the development and evaluation of digital imaging technologies and image analysis software as well as their translation into safe and effective clinical services.
Disclaimer

- The CAP does not permit reproduction of any substantial portion of the material in this Webinar without its written authorization. The CAP hereby authorizes attendees of the CAP Webinar to use the PDF presentation solely for educational purposes within their own institutions. The CAP prohibits use of the material in the Webinar – and any unauthorized use of the CAP’s name or logo – in connection with promotional efforts by marketers of laboratory equipment, reagents, materials, or services.
Disclaimer

- Opinions expressed by the speaker are the speaker’s own and do not necessarily reflect an endorsement by the CAP of any organizations, equipment, reagents, materials, or services used by participating laboratories.
Disclaimer

- I have no conflicts of interests related to this presentation.
Outline

1. Limitations of current approach to machine learning model development and deployment
2. Proposed solution to develop turnkey models that can be readily deployed for clinical application
3. Proof-of-concept use case
Ingredients of a machine learning project

Data
- Cohort creation
- Data curation
- Data annotation

Objective
- \( \arg \max_{\theta \in \Theta} L(X; \theta) \)
- Problem definition
- Loss function
- Evaluation criteria

Algorithm
- Model training
- Model testing
- Model inference

Software
- Libraries
- Applications
- Platforms

Hardware
- Compute
- Storage
- Network

Pathologists ↔ Data Scientists ↔ Software Engineers

© College of American Pathologists 10 June 2021
The ideal machine learning project

Concept
- Clinical Need
  - Data Science
- Data Set
  - Model Development
- Research
  - Open-source Software Tools

Prototype
- Publication
  - Model Development
- Model
  - Model Deployment
- Clinical Need
  - Publication
  - Prototype
  - Concept

Product
- Clinical Service
  - Medical Device
  - Clinical Practice
  - Commercial Software Systems
The computational pathology gap

**Development:** Data Scientists

**Deployment:** Software Engineers

https://cloud.google.com/healthcare
Data communication in machine learning

Model Development

Establish ground truth

Use ground truth for model training

Model Deployment

Interpret model outputs

Render predictions upon model inference

Model Development

Image Annotation

Establish ground truth

Use ground truth for model training

Model Deployment

Decision Support

Interpret model outputs

Render predictions upon model inference

Model Training

Image Display System

Labels Images

Image Management System

Labels Images

Image Analysis System

Labels Images

© College of American Pathologists

10 June 2021
Bridging the data communication gap

Goal

Develop turnkey, deployment-ready models (**plug & play**)

Proposed solution

- Apply **consistent data management** throughout the model development and deployment process
- Rely on a **standard interface** to achieve **interoperability**
“The ability of two or more systems or components to exchange information and to use the information that has been exchanged.”

Institute of Electrical and Electronics Engineers (IEEE) Standard Computer Dictionary
“Secure exchange of electronic health information with, and use of electronic health information from, other health information technology without special effort on the part of the user.”

21st Century Cures Act, Section 4003 (A)
Outline

1. Limitations of current approach to machine learning model development and deployment
2. Proposed solution to develop turnkey models that can be readily deployed for clinical application
3. Proof-of-concept use case
Digital Imaging and Communications in Medicine

Herrmann et al. 2018
Image annotations and model outputs in DICOM

Comprehensive 3D SR
- Patient
- Study
- Series
- Equipment

Document
- Name
  - CODE
  - SCOORD3D
  - Value
- Name
  - CODE
  - Value
- Name
  - CODE
  - NUM
  - Value
- Name
  - CODE
  - IMAGE
  - Value

VL Whole Slide Microscopy Image
- Patient
- Study
- Series
- Equipment
- Specimen
- Frame of Reference
- Image
  - Frame
  - Frame
  - Frame
  - Frame

Segmentation
- Patient
- Study
- Series
- Equipment
- Specimen
- Frame of Reference
- Image
  - Frame
  - Frame
  - Frame
  - Frame

Bridge et al. In preparation
DICOM-enabled machine learning platform

Google Cloud
Cloud Healthcare API

© College of American Pathologists
Demo of *slim* viewer with Google Healthcare API
The future of bioimaging research in the cloud

Image Data Repository
- Radiological images
- Slide microscopy images
- Image annotations
- Image-derived measurements

Image Management
DICOMweb interface of the Google Healthcare API

https://datacommons.cancer.gov
Outline

1. Limitations of current approach to machine learning model development and deployment

2. Proposed solution to develop turnkey models that can be readily deployed for clinical application

3. Proof-of-concept use case
Lung cancer detection in whole slide images

Data-efficient and weakly supervised computational pathology on whole-slide images

Ming Y. Lu\textsuperscript{1,2,3}, Drew F. K. Williamson\textsuperscript{1,5}, Tiffany Y. Chen\textsuperscript{1,5}, Richard J. Chen\textsuperscript{1,4}, Matteo Barbieri\textsuperscript{1,2} and Faisal Mahmood\textsuperscript{1,2,3,5}

Deep-learning methods for computational pathology require either manual annotation of gigapixel whole-slide images (WSIs) or large datasets of WSIs with slide-level labels and typically suffer from poor domain adaptation and interpretability. Here, we report an interpretable weakly supervised deep-learning method for data-efficient WSI processing and learning that only requires slide-level labels. The method, which we named clustering-constrained-attention multiple-instance learning (CLAM), uses attention-based learning to identify subregions of high diagnostic value to accurately classify whole slides and instance-level clustering over the identified representative regions to constrain and refine the feature space. By applying CLAM to the subtyping of renal cell carcinoma and non-small-cell lung cancer as well as the detection of lymph node metastasis, we show that it can be used to localize well-known morphological features on WSIs without the need for spatial labels, that it outperforms standard weakly supervised classification algorithms and that it is adaptable to independent test cohorts, smartphone microscopy and varying tissue content.
Collections of whole slide images in DICOM format

https://portal.imaging.datacommons.cancer.gov
Collections from The Cancer Imaging Archive

TCGA (n=3114)
- LUAD (n=1541)
- LSCC (n=1573)

CPTAC (n=790)
- LUAD (n=496)
- LSCC (n=294)
DICOM-based model training and inference

Series of SM Images

SM Image

SR Document

TID 1410 Planar ROI Measurements and Qualitative Evaluations

<table>
<thead>
<tr>
<th>Finding [DCM, 121071]</th>
<th>Neoplasm [SCT, 108389006]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding Site [SCT, 363886007]</td>
<td>Lung [SCT, 36607008]</td>
</tr>
<tr>
<td>Morphology [SCT, 116670008]</td>
<td>Adenocarcinoma, NOS [ICD03, 8140/3]</td>
</tr>
<tr>
<td>Topography [SCT, 116677004]</td>
<td>Bronchus and lung [C10C, C34]</td>
</tr>
<tr>
<td>Image Region [DCM, 111030]</td>
<td>POLYGON $x_1 y_1 z_1 x_2 y_2 z_2 ... x_6 y_6 z_6$</td>
</tr>
<tr>
<td>Area [SCT, 42798000]</td>
<td>12.07 Square millimeter [UCUM, mm2]</td>
</tr>
</tbody>
</table>

Bridge et al. In preparation
Model outputs for selected tumor tissue sections

LUAD
LUAD
LUSC
LUSC
Model outputs for selected tumor tissue section

- Tissue
- LUAD: 4.16%
- LUSC: 54.25%
- NSCLC: 58.41%
Exchange of model outputs in DICOM format

Region of Interests (ROI)
Frame of reference

Qualitative Evaluations
Coded concepts

Measurements
Units

https://whobluebooks.iarc.fr/
Encoding of model outputs in DICOM

```python
import numpy as np
import highdicom as hd
from pydicom.irpc.codes import import_codes

# Encode bounding box enclosing the tumor region using 3D spatial coordinates
region_item = hd.sr.RegionItem3D(
    graphic_type=hd.sr.GraphicTypeValues3D.POLYGON,
    graphic_data=np.array([[(45., 34., 0.), (40., 34., 0.), (40., 42., 0.), (45., 42., 0.), (45., 34., 0.)]],
    frame_of_reference_uid="1.2.3.4"
)

# Encode a qualitative evaluation describing the "morphology" as "adenocarcinoma"
morphology_item = hd.sr.CodedConceptItem(
    name=hd.sr.CodedConcept(
        value=hd.sr.CodedConcept(
            value="8140/3",
            scheme_designator="ICD9",
            meaning="Adenocarcinoma, NOS"
        )
    )
)

# Encode a measurement with name "area" and of unit "square millimeter"
area_item = hd.sr.NumContentItem(
    name=hd.sr.NumContentItem(
        value=12.07,
        unit=hd.sr.CodedConcept(
            value=hd.sr.CodedConcept(
                value="1917/0",
                scheme_designator="UCUM",
                meaning="Square millimeter"
            )
        )
    )
)

import highdicom as hd
from highdicom.dataset import Dataset
from highdicom.sr.codes import import_codes

def load_image() -> Dataset:
    ...

source_image = load_image()

# Describe the device that is reporting the observation
observation_context = hd.sr.ObservationContext(
    observer_device_context=hd.sr.ObserverContext(
        observer_type=hd.sr.ObservationContext(
            observer_id=hd.sr.ObservationContext(
                identifier=hd.sr.ObservationContext().identifier,
                name="GPU01"
            )
        )
    )
)

# Describe the region of interest finding and associated measurements and evaluations
roi_description = hd.sr.PlanarROIMeasurementsAndQualitativeEvaluations(
    finding_type=hd.sr.FindingType(Anterior, location=hd.sr.CodedConcept(
        value=hd.sr.CodedConcept(
            value="1917/0",
            scheme_designator="UCUM",
            meaning="Square millimeter"
        )
    )
)

# Construct a measurement report from the ROI(s)
measurement_report = hd.sr.MeasurementReport(
    observation_context=observation_context,
    procedure_reported=hd.sr.CodedConcept(
        value=hd.sr.CodedConcept(
            value="1917/0",
            scheme_designator="UCUM",
            meaning="Square millimeter"
        )
    )
)

# Construct an SR document containing the measurement report
report_document = hd.sr.Comprehensive3DSR(
    evidence=source_image,
    content=measurement_report)
```

© College of American Pathologists 10 June 2021 28
Review of model outputs in *slim* viewer
Open-source software available on GitHub

MGH Computational Pathology
Department of Pathology, Massachusetts General Hospital
Boston, MA

Pinned repositories

- **slim**
  Interoperable slide microscopy viewer for thin clients
  - TypeScript
  - 7 dependencies
  - 4 contributors

- **dicom-microscopy-viewer**
  Web-based viewer for DICOM Visible Light Whole Slide Microscopy Images
  - JavaScript
  - 46 dependencies
  - 13 contributors

- **dicomweb-client**
  Python client for DICOMweb RESTful services
  - Python
  - 66 dependencies
  - 20 contributors
Participation in DICOM WG-26 Connectathon
DICOM WG-26 Pathology

WG-26: Pathology

Secretariat (U.S)  
College of American Pathologists (CAP)  
Kevin Schap, wg26chairs@dicomstandard.org

Secretariat (Europe)  
Sociedad Española de Informática de la Salud (SEIS)  
Marcial García Rojo, wg26chairs@dicomstandard.org

Co-Chairs  
Markus Herrmann, Massachusetts General Hospital  
Mikael Wintell, Västra Götalands Regionen  
wg26chairs@dicomstandard.org

Last strategy update  
2019-10-02

Minutes  
WG-26 minutes

Scope:  
- WG26's scope is to support and develop the DICOM Standard so the Pathology Domain can handle their whole slide images and also the macros when it comes to produce, store and communicate. We also want to make it possible for the domain to be a part of the whole healthcare process by open up those system when it comes to interoperability using global Standards to create workflows.
DICOM WG-26 Hackathon

Pathology Hackathon

Develop **interoperable computational tools** for whole slide image annotation and analysis

**May 1st - June 18th 2021** (finale at ECDP 2021 congress)

https://github.com/DICOMWG26AnnotationAHG/hackathon
Collaborative development of open standards

DICOM WG-26 WSI Annotation Hackathon

A virtual Hackathon for collaborative development and testing of software for the generation, exchange, and use of whole slide image (WSI) annotations in DICOM SR, SEG or Sup 222 format, in which anyone (vendors, academics and individuals) can participate freely and openly.

The event will be held virtually over the internet and coordinated via this repository. It will start May 1st 2021 and end with a grand finale at the European Congress on Digital Pathology (ECDP) June 15th-18th, 2021. There will be opportunities at the ECDP conference to demo software and present results to an international audience of experts in digital and computational pathology.
Integrating the Healthcare Enterprise (IHE)

IHE PaLM

Digital Pathology Image Acquisition

Dash et al. 2021
Acknowledgements

MGH Pathology
- Christopher Gorman
- Aidan Stein
- Veronica Klepeis
- Jochen Lennerz
- Lida Harriri
- Long Le
- John Iafrate
- David Louis

MGH Radiology / MGB CCDS
- Christopher Bridge
- Sean Doyle
- Jayashree Kalpathy-Cramer

BWH Radiology
- Andrey Fedorov
- Ron Kikinis

Isomics
- Steven Pieper

Radical Imaging
- Igor Octaviano
- Davide Punzo
- Eric Ziegler

Google
- Jason Klotzer
- Craig Mermel

DICOM WG-26
- David Clunie (Pixelmed)
- Gábor Garányi (3DHistech)
- Steve Moore (WashU)
- Norman Zerbe (Charité)
- Joachim Schmidt (Roche)
- Kevin Schap (CAP)

IHE PaLM
- Rash Dash (Duke)
- Mary Kennedy (CAP)

National Cancer Institute
- Ulrike Wagner
- Keyvan Farahani
Resources of Digital and Computational Pathology Committee

• List resources
  o SPECs
  o Resources Guides
  o Topic Center Pages & AI pages
  o AI@CAP.ORG email address
THANK YOU!

Thank you for attending our webinar “Realizing Artificial Intelligence in Pathology – Streamlining Machine Learning Model Development and Clinical Deployment” by Markus D. Herrmann, M.D., Ph.D.

For comments about this webinar or suggestions for upcoming webinars, contact Al@cap.org

NOTE: There is no CME/CE credit available for today’s complimentary webinar. The pdf of the presentation will be sent out in about 1 week.