**CHALLENGES AND STRATEGIES FOR DEVELOPING ROBUST AI APPLICATIONS IN ONCOLOGY**

Daniel L. Rubin, MD, MS

Professor of Biomedical Data Science, Radiology, Medicine (Biomedical Informatics), Computer Science (courtesy) and Ophthalmology (courtesy)

Laboratory of Quantitative Imaging and AI

Stanford University

---

**Image assessments of cancer lesions are the basis of evaluating treatment response**

- Measurements are made on images
- Recorded as a separate process (spreadsheet, dictated report, etc)
- Disconnected from the image(s)

---

**Problem: Inter-reader variation:**

Sufficiency of information in 167 imaging studies and reports

<table>
<thead>
<tr>
<th>No of Observations</th>
<th>Baseline (55)</th>
<th>Follow-up (112)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesion Described</td>
<td>39 (71%)</td>
<td>43 (38%)</td>
</tr>
<tr>
<td>Longest Diameter</td>
<td>30 (55%)</td>
<td>31 (28%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image Markup</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion Identified</td>
<td>40 (73%)</td>
<td>78 (70%)</td>
</tr>
<tr>
<td>Longest Diameter</td>
<td>27 (50%)</td>
<td>29 (26%)</td>
</tr>
</tbody>
</table>

Radiologist not aware of measurable disease being tracked by oncologists

---

**Opportunities for AI in cancer imaging**

- Lesion detection
- Lesion segmentation
- Diagnosis
- Treatment selection
- Response assessment
- Clinical prediction (of treatment response or future disease)

Critical for Drug Evaluation

Active research area

---

**Acknowledgements**

Students, Post-docs, Residents, Staff, and Collaborators

- Bao Do
- Christopher Re
- Sandy Napel
- Chris Beaulieu
- Ozge Yurtsever
- Darvin Yi
- Anuj Pareek
- Jared Dunnmon
- Anuj Pareek
- Siyi Tang
- Maha Alkim
- David Liang
- Niranjan Balasubramanian
- Emel Akkim
- Michael Gensheimer
- Nandita Bhaskar
- Ted Leng
- Joelle Hallak
- Luis de Sisternes
- Ozge Yurtsever
- Bede Do
- Sandy Napel
- Chris Beaulieu
- Ami Leung
- Naras Balasubramanian
- Chris Beaulieu
- Ann Leung
- Phil Lavori
- Jarred Dunnmon
- Mete Akdogan
- Anna Leung
- Christopher Re
- Sandy Napel
- Chris Beaulieu

Funding Support

- NCI grants:
  - U01CA142555
  - 1U01CA190214
  - 1U01CA187947
  - 1U01CA242879
- NVIDIA Academic Hardware Grant Program

- Stanford Philips and GE BlueSky

---

**Lesion measurements are basis of patient response and cohort treatment efficacy**

*Sum of Maximum Lesion Diameters Over Time*

*Problem:*

These analyses are currently hand-generated in a cumbersome, error-prone workflow, subject to inter-reader variation.

---

Image assessments of cancer lesions are the basis of evaluating treatment response

*Measurements are made on images*

*Recorded as a separate process (spreadsheet, dictated report, etc)*

*Disconnected from the image(s)*

---

**No of Observations**

<table>
<thead>
<tr>
<th>Baseline (55)</th>
<th>Follow-up (112)</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 (71%)</td>
<td>43 (38%)</td>
</tr>
<tr>
<td>30 (55%)</td>
<td>31 (28%)</td>
</tr>
<tr>
<td>40 (73%)</td>
<td>78 (70%)</td>
</tr>
<tr>
<td>27 (50%)</td>
<td>29 (26%)</td>
</tr>
</tbody>
</table>

---

**Opportunities for AI in cancer imaging**

- Lesion detection
- Lesion segmentation
- Diagnosis
- Treatment selection
- Response assessment
- Clinical prediction (of treatment response or future disease)

---

Copyright © Daniel Rubin 2019
Detection and segmentation: General fully connected networks

Detection/segmentation are pixel-based classification tasks

There are challenges to building robust AI models

- Data among institutions varies
  - Geographic variations in patient populations
  - Differences in imaging parameters
  - Differences in vendor equipment
- Robust AI models require **large amounts of labeled training** data in order to generalize
- Difficult/costly to acquire large amounts of data
- There are tremendous amounts of historical data across institutions that could be leveraged

AI development (and data) is siloed

Overcoming barriers to data sharing

- **Bring the model to the data** instead of bringing the data (centralized) to the model
- **Distributed computation** of training deep learning models ("distributed learning")
**Alternative models for training distributed deep learning models**

<table>
<thead>
<tr>
<th>A</th>
<th>Centrally hosted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution 1</td>
<td></td>
</tr>
<tr>
<td>Institution 2</td>
<td></td>
</tr>
<tr>
<td>Institution 3</td>
<td></td>
</tr>
<tr>
<td>Institution 4</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Ensemble single institution</td>
</tr>
<tr>
<td>Institution 1</td>
<td></td>
</tr>
<tr>
<td>Institution 2</td>
<td></td>
</tr>
<tr>
<td>Institution 3</td>
<td></td>
</tr>
<tr>
<td>Institution 4</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Single weight transfer</td>
</tr>
<tr>
<td>Institution 1</td>
<td></td>
</tr>
<tr>
<td>Institution 3</td>
<td></td>
</tr>
<tr>
<td>Institution 4</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Cyclic weight transfer (CWT)</td>
</tr>
<tr>
<td>Institution 1</td>
<td></td>
</tr>
<tr>
<td>Institution 3</td>
<td></td>
</tr>
<tr>
<td>Institution 4</td>
<td></td>
</tr>
</tbody>
</table>

**CWT has similar performance to centrally-hosted training**

![Graph showing comparison between Centrally Hosted and CWT](chart)

**Results based on having 4 institutions**

**Challenges to making distributed learning work**

*Distributed model performance may be inferior to centrally hosted performance*

- **Heterogeneity in data** across institutions
  - Different patient populations (sample sizes, label distribution)
  - Differences in manifestation of disease
  - Different scanning hardware/parameters
  - Differences in image resolution
  - Differences in image quality
- **Differences in computing hardware** (GPU/CPU) among institutions
- **Differences in network bandwidth** among institutions

**Variability in sample size and label distribution (NIH Chest X-ray dataset)**

- 4 institutions, GoogleNet for classification
- Each point is average across 10 runs
- We generate various institution splits such that there are different levels of variance in sample size and label distribution

**Overcoming challenges of variability in data (NIH Chest X-ray dataset)**

- 4 institutions, GoogleNet for classification
- Each point is average across 10 runs
- We generate various institution splits such that there are different levels of variance in sample size and label distribution

**Importance of sample size**

- **Centrally hosted data**
  - 6000 patients
  - Each site has 300 cases
  - Accuracy increases with number of collaborating institutions—amount of data (each site has 300 cases)

**Copyright © 2019 Daniel L. Rubin**
The future: A collaborative AI ecosystem

• Each site specifies a dataset they wish to make available for building AI
• “Pop-up” collaborations to build AI models using distributed learning
• Full control over data and data use

Summary

• We need AI in cancer imaging to reduce variations in assessment of lesions
• Key AI methods are automated lesion detection and segmentation
• Building robust AI models requires much data
• Data from multiple institutions can be leveraged through distributed learning
  — Heterogeneities among sites is a challenge
  — Optimizations in distributed computational approaches can overcome challenges

Thank you.

Contact info:
dlrubin@stanford.edu