AI For Healthcare

Pranav Rajpurkar
PhD Candidate, Computer Science
Stanford ML Group
Diabetic Retinopathy from retinal fundus photos (Gulshan et al., 2016)
Melanomas from photos of skin lesions (Esteva et al., 2017)
Lymph node metastases from H&E slides (Bejnordi et al., 2018)
Arrhythmias from ECG signals (Hannun & Rajpurkar et al., 2019)
What are **challenges** for building and deploying AI for medical image interpretation?

What are **opportunities** for making AI medical image interpretation routine part of clinical care?
Challenges for AI in Medical Image Interpretation

Data

Deployment
Data Challenges

● Prevalence
● Noisy Labels
● Evaluation
CheXNeXt

Rajpurkar & Irvin et al., PLOS Medicine, 2018
Data Challenges

- Prevalence
- Noisy Labels
- Evaluation
Prevalence Challenge

Normal
Normal
Nodule
Normal
Training Data as seen in real world

P1 Normal
P2 Normal
P3 Normal
P4 Nodule
P5 Normal
P6 Normal
P7 Nodule
P8 Normal

Learning Algorithm

Poor Test Set Performance

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003
Training Data as seen in real world

\[
L(X, y) = -y \log p(Y = 1|X) - (1 - y) \log p(Y = 0|X)
\]

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003

Cost Sensitive Learning

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>Nodule</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>Nodule</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>Normal</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total Contribution

<table>
<thead>
<tr>
<th>Normal</th>
<th>Nodule</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003
### Cost Sensitive Learning

<table>
<thead>
<tr>
<th>Patient</th>
<th>Diagnosis</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P2</td>
<td>Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P3</td>
<td>Normal</td>
<td>2/8</td>
</tr>
<tr>
<td><strong>P4</strong></td>
<td><strong>Nodule</strong></td>
<td><strong>6/8</strong></td>
</tr>
<tr>
<td>P5</td>
<td>Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P6</td>
<td>Normal</td>
<td>2/8</td>
</tr>
<tr>
<td><strong>P7</strong></td>
<td><strong>Nodule</strong></td>
<td><strong>6/8</strong></td>
</tr>
<tr>
<td>P8</td>
<td>Normal</td>
<td>2/8</td>
</tr>
</tbody>
</table>

\[
L(X, y) = -w_+ \cdot y \log p(Y = 1 | X) - w_- \cdot (1 - y) \log p(Y = 0 | X)
\]

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003
**Cost Sensitive Learning**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P2 Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P3 Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P4 Nodule</td>
<td>6/8</td>
</tr>
<tr>
<td>P5 Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P6 Normal</td>
<td>2/8</td>
</tr>
<tr>
<td>P7 Nodule</td>
<td>6/8</td>
</tr>
<tr>
<td>P8 Normal</td>
<td>2/8</td>
</tr>
</tbody>
</table>

**Total Contribution**

- **Normal**
  \[
  6 \times \frac{2}{8} = \frac{12}{8}
  \]

- **Nodule**
  \[
  2 \times \frac{6}{8} = \frac{12}{8}
  \]

---

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003

Cost Sensitive Learning

P1 Normal 2/8
P2 Normal 2/8
P3 Normal 2/8
P4 Nodule 6/8
P5 Normal 2/8
P6 Normal 2/8
P7 Nodule 6/8
P8 Normal 2/8

Learning Algorithm

Good Test Set Performance

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003
### Cost Sensitive Learning

<table>
<thead>
<tr>
<th>Pathology</th>
<th>AUC of algorithm trained on original training set (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelectasis</td>
<td>0.843 (0.803,0.878)</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.827 (0.784,0.867)</td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.860 (0.819,0.896)</td>
</tr>
<tr>
<td>Edema</td>
<td>0.933 (0.901,0.96)</td>
</tr>
<tr>
<td>Effusion</td>
<td>0.901 (0.869,0.929)</td>
</tr>
<tr>
<td>Emphysema</td>
<td>0.790 (0.709,0.862)</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>0.809 (0.705,0.9)</td>
</tr>
<tr>
<td>Hernia</td>
<td>0.782 (0.704,0.852)</td>
</tr>
<tr>
<td>Infiltration</td>
<td>0.710 (0.636,0.778)</td>
</tr>
<tr>
<td>Mass</td>
<td>0.880 (0.828,0.926)</td>
</tr>
<tr>
<td>Nodule</td>
<td>0.882 (0.842,0.92)</td>
</tr>
<tr>
<td>Pleural Thickenning</td>
<td>0.783 (0.731,0.834)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.774 (0.678,0.858)</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.917 (0.868,0.956)</td>
</tr>
</tbody>
</table>

Weiss et al. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. 2003
Rajpurkar & Irvin et al., PLOS Medicine, 2018
Data Challenges

- Prevalence
- Noisy Labels
- Evaluation
Noisy Labels Challenge
Noisy Labels

Oakden-Rayner. Exploring the ChestXray14 dataset: problems. 2017
Keeping the Noise

Noisy Label

Classifier

## Keeping the Noise

Noisy Label

Classifier

<table>
<thead>
<tr>
<th>Pathology</th>
<th>AUC of algorithm trained on original training set (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelectasis</td>
<td>0.843 (0.803,0.878)</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.827 (0.784,0.867)</td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.860 (0.819,0.896)</td>
</tr>
<tr>
<td>Edema</td>
<td>0.933 (0.901,0.96)</td>
</tr>
<tr>
<td>Effusion</td>
<td>0.901 (0.869,0.929)</td>
</tr>
<tr>
<td>Emphysema</td>
<td>0.790 (0.709,0.862)</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>0.809 (0.705,0.9)</td>
</tr>
<tr>
<td>Hernia</td>
<td>0.782 (0.704,0.852)</td>
</tr>
<tr>
<td>Infiltration</td>
<td>0.710 (0.636,0.778)</td>
</tr>
<tr>
<td>Mass</td>
<td>0.880 (0.828,0.926)</td>
</tr>
<tr>
<td>Nodule</td>
<td>0.882 (0.842,0.92)</td>
</tr>
<tr>
<td>Pleural Thickening</td>
<td>0.783 (0.731,0.834)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.774 (0.678,0.858)</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.917 (0.868,0.956)</td>
</tr>
</tbody>
</table>

Rajpurkar & Irvin et al., PLOS Medicine, 2018
Keeping the Noise

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Prevalence (examples)</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effusion</td>
<td>11,000</td>
<td>0.901</td>
</tr>
<tr>
<td>Hernia</td>
<td>110</td>
<td>0.851</td>
</tr>
</tbody>
</table>

Diabetic Retinopathy from retinal fundus photos (Gulshan et al., 2016)
Melanomas from photos of skin lesions (Esteva et al., 2017)
Rajpurkar & Irvin et al., PLOS Medicine, 2018
Correcting Noise via Self-Training

Noisy Label

Train

Classifier

Label

Corrected Label

Train

Classifier

Rajpurkar & Irvin et al., PLOS Medicine, 2018
Correcting Noise

<table>
<thead>
<tr>
<th>Pathology</th>
<th>AUC of algorithm trained on original training set (95% CI)</th>
<th>AUC of algorithm trained on relabeled training set (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelectasis</td>
<td>0.843 (0.803,0.878)</td>
<td>0.862 (0.825,0.895)</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.827 (0.784,0.867)</td>
<td>0.831 (0.79,0.87)</td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.860 (0.819,0.896)</td>
<td>0.893 (0.859,0.924)</td>
</tr>
<tr>
<td>Edema</td>
<td>0.933 (0.901,0.96)</td>
<td>0.924 (0.886,0.955)</td>
</tr>
<tr>
<td>Effusion</td>
<td>0.901 (0.869,0.929)</td>
<td>0.901 (0.868,0.93)</td>
</tr>
<tr>
<td>Emphysema</td>
<td>0.790 (0.709,0.862)</td>
<td>0.704 (0.567,0.833)</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>0.809 (0.705,0.9)</td>
<td>0.806 (0.719,0.884)</td>
</tr>
<tr>
<td>Hernia</td>
<td>0.782 (0.704,0.852)</td>
<td>0.851 (0.785,0.909)</td>
</tr>
<tr>
<td>Infiltration</td>
<td>0.710 (0.636,0.778)</td>
<td>0.721 (0.651,0.786)</td>
</tr>
<tr>
<td>Mass</td>
<td>0.880 (0.828,0.926)</td>
<td>0.909 (0.864,0.948)</td>
</tr>
<tr>
<td>Nodule</td>
<td>0.882 (0.842,0.92)</td>
<td>0.894 (0.853,0.93)</td>
</tr>
<tr>
<td>Pleural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickening</td>
<td>0.783 (0.731,0.834)</td>
<td>0.798 (0.744,0.849)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.774 (0.678,0.858)</td>
<td>0.851 (0.781,0.911)</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.917 (0.868,0.956)</td>
<td>0.944 (0.915,0.969)</td>
</tr>
</tbody>
</table>

Rajpurkar & Irvin et al., PLOS Medicine, 2018
Data Challenges

● Prevalence
● Noisy Labels
● Evaluation
Evaluation Challenge
Evaluation Challenge

Noisy Ground Truth Label

Strong Ground Truth

Train

Classifier
Diabetic Retinopathy from retinal fundus photos (Gulshan et al., 2016)
Melanomas from photos of skin lesions (Esteva et al., 2017)
Rajpurkar & Irvin et al., PLOS Medicine, 2018
Test Set

Ground Truth

AI

Reader

v.s.
Model Comparable to Radiologists

Except when small number of examples (e.g. Hernia)
Recent study comparing several chest x-ray models

CheXNeXt training strategy competitive
Data Challenges

- Prevalence
- Noisy Labels
- Evaluation

Cost-Sensitive Learning

Weak-Supervision at Scale

Strong Ground Truth Collection
Data Challenges

- Prevalence
- Noisy Labels
- Evaluation

Design Datasets
Design datasets that solve data challenges

200k Chest X-Rays
1,370 Knee MRI exams
40k Bone x-rays

Work with Dr. Matt Lungren, Dr. Curt Langlotz and many others at the Stanford Med School

Irvin & Rajpurkar et al. 2019, AAAI; Bien & Rajpurkar et al, PLOS Medicine; Rajpurkar & Irvin et al. 2018, MIDL
Designing to Tackle Prevalence

200k Chest X-Rays

Irvin & Rajpurkar et al. 2019, AAAI
Designing to Tackle Noisy Labels

1. *unremarkable* cardiome diastinal silhouette

2. diffuse *reticular pattern*, which can be seen with an atypical *infection* or chronic fibrotic change. *no focal consolidation*.

3. *no pleural effusion* or *pneumothorax*

4. mild degenerative changes in the lumbar spine and old right rib *fractures*.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Labeler Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finding</td>
<td>0</td>
</tr>
<tr>
<td>Enlarged Cardiom.</td>
<td></td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td></td>
</tr>
<tr>
<td>Lung Opacity</td>
<td>1</td>
</tr>
<tr>
<td>Lung Lesion</td>
<td></td>
</tr>
<tr>
<td>Edema</td>
<td></td>
</tr>
<tr>
<td>Consolidation</td>
<td>0</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>u</td>
</tr>
<tr>
<td>Atelectasis</td>
<td></td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0</td>
</tr>
<tr>
<td>Pleural Effusion</td>
<td>0</td>
</tr>
<tr>
<td>Pleural Other</td>
<td></td>
</tr>
<tr>
<td>Fracture</td>
<td>1</td>
</tr>
<tr>
<td>Support Devices</td>
<td></td>
</tr>
</tbody>
</table>
Designing to Tackle Noisy Labels

<table>
<thead>
<tr>
<th>Category</th>
<th>Mention F1 NIH</th>
<th>Mention F1 Ours</th>
<th>Negation F1 NIH</th>
<th>Negation F1 Ours</th>
<th>Uncertain F1 NIH</th>
<th>Uncertain F1 Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelectasis</td>
<td>0.976</td>
<td><strong>0.998</strong></td>
<td>0.526</td>
<td><strong>0.833</strong></td>
<td>0.661</td>
<td><strong>0.936</strong></td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.647</td>
<td><strong>0.973</strong></td>
<td>0.000</td>
<td><strong>0.909</strong></td>
<td>0.211</td>
<td><strong>0.727</strong></td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.996</td>
<td><strong>0.999</strong></td>
<td>0.879</td>
<td><strong>0.981</strong></td>
<td>0.438</td>
<td><strong>0.924</strong></td>
</tr>
<tr>
<td>Edema</td>
<td>0.978</td>
<td><strong>0.993</strong></td>
<td>0.873</td>
<td><strong>0.962</strong></td>
<td>0.535</td>
<td><strong>0.796</strong></td>
</tr>
<tr>
<td>Pleural Effusion</td>
<td>0.985</td>
<td><strong>0.996</strong></td>
<td>0.951</td>
<td><strong>0.971</strong></td>
<td>0.553</td>
<td><strong>0.707</strong></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.660</td>
<td><strong>0.992</strong></td>
<td>0.703</td>
<td><strong>0.750</strong></td>
<td>0.250</td>
<td><strong>0.817</strong></td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.993</td>
<td><strong>1.000</strong></td>
<td>0.971</td>
<td><strong>0.977</strong></td>
<td>0.167</td>
<td><strong>0.762</strong></td>
</tr>
</tbody>
</table>
## Designing to Tackle Noisy Labels

### MIMIC-CXR Database

<table>
<thead>
<tr>
<th>Category</th>
<th>Mention</th>
<th></th>
<th>Negation</th>
<th></th>
<th>Uncertainty</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NegBio</td>
<td>CheXpert</td>
<td>NegBio</td>
<td>CheXpert</td>
<td>NegBio</td>
<td>CheXpert</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>0.930</td>
<td>0.998</td>
<td>0.727</td>
<td>0.400</td>
<td>0.379</td>
<td>0.835</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.596</td>
<td>0.954</td>
<td>0.043</td>
<td>0.830</td>
<td>0.000</td>
<td>0.333</td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.966</td>
<td>0.986</td>
<td>0.917</td>
<td>0.958</td>
<td>0.235</td>
<td>0.486</td>
</tr>
<tr>
<td>Edema</td>
<td>0.855</td>
<td>0.996</td>
<td>0.701</td>
<td>0.878</td>
<td>0.214</td>
<td>0.742</td>
</tr>
<tr>
<td>Pleural Effusion</td>
<td>0.971</td>
<td>0.987</td>
<td>0.873</td>
<td>0.947</td>
<td>0.368</td>
<td>0.500</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.836</td>
<td>0.981</td>
<td>0.750</td>
<td>0.785</td>
<td>0.388</td>
<td>0.674</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.983</td>
<td>0.998</td>
<td>0.951</td>
<td>0.948</td>
<td>0.182</td>
<td>0.286</td>
</tr>
</tbody>
</table>

*Johnson et al. 2019. MIMIC-CXR. 2019*
Designing to Tackle Noisy Labels

NegBio User Guide

Run the pipeline step-by-step

The step-by-step pipeline generates all intermediate documents. You can easily rerun one step if it makes errors. The whole steps are

1. **text2bioc** combines text into a BioC XML file.
2. **normalize** removes noisy text such as [**Patterns**].
3. **section_split** splits the report into sections based on titles at `patterns/section_titles.txt`.
4. **ssplit** splits text into sentences.
5. Named entity recognition
   - **dner_mm** detects UMLS concepts using MetaMap.
   - **dner_chexpert** detects concepts using the CheXpert vocabularies at `negbio/chexpert/phrases`.
6. **parse** parses sentence using the Bllip parser.
7. **ptb2ud** converts the parse tree to universal dependencies using Stanford converter.
8. Negation detection
   - **neg** detects negative and uncertain findings.
   - **neg_chexpert** detects positive, negative and uncertain findings (recommended).

Designing to Tackle Evaluation

Irvin & Rajpurkar et al. 2019, AAAI
### Designing to Tackle Evaluation

<table>
<thead>
<tr>
<th>Rank</th>
<th>Date</th>
<th>Model</th>
<th>AUC</th>
<th>Num Rads Below Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sep 01, 2019</td>
<td>Hierarchical-Learning-V1 (ensemble) Vingroup Big Data Institute</td>
<td>0.930</td>
<td>2.6</td>
</tr>
<tr>
<td>2</td>
<td>Oct 10, 2019</td>
<td>yww</td>
<td>0.929</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>Sep 09, 2019</td>
<td>Multi-Stage-Learning-CNN-V3 (ensemble) VINBDI Medical Imaging Team</td>
<td>0.928</td>
<td>2.6</td>
</tr>
<tr>
<td>4</td>
<td>Sep 19, 2019</td>
<td>SenseXDR ensemble</td>
<td>0.927</td>
<td>2.6</td>
</tr>
<tr>
<td>5</td>
<td>Sep 18, 2019</td>
<td>ihil (ensemble) UESTC</td>
<td>0.927</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Data Challenges

- Prevalence
- Noisy Labels
- Evaluation

Design Datasets
Challenges for AI in Medical Image Interpretation

Data

Deployment
Deployment Challenges

- *Algorithm Translation*
- *Clinical Impact*
Deployment Challenges

- **Algorithm Translation**
- **Clinical Impact**
Algorithm Translation Challenge

Translation Engineering

Model

- Atelectasis
- Cardiomegaly
- Consolidation
- Edema
- Effusion
- Emphysema
- Fibrosis
- Hernia
- Infiltration
- Mass
- Nodule
- Pleural
- Thickening
- Pneumonia
- Pneumothorax
Engineering for Translation: XRay4All

With Amir Kiani, Dr. Matt Lungren, and others, Stanford ML Group
Engineering for Translation: XRay4All
XRay4All: Live demo
Deployment Challenges

- *Algorithm Translation*
- *Clinical Impact*
Clinical Impact Challenge

Model

Performance

→

Model

Clinician

Performance
HeadXNet

Park & Chute & Rajpurkar et al., JAMA Network Open. 2019
Pneumonia

2D CNN

3D CNN

Pneumonia
HeadXNet takes in a series of CTA images and outputs segmentation for each voxel.
High Sensitivity Model for Detecting Aneurysms

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Accuracy (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinicians</strong></td>
<td>0.831 (0.794, 0.862)</td>
<td>0.960 (0.937, 0.974)</td>
<td>0.893 (0.782, 0.912)</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>0.949 (0.860, 0.963)</td>
<td>0.661 (0.530, 0.771)</td>
<td>0.809 (0.727, 0.870)</td>
</tr>
</tbody>
</table>
A. Crossover Study Design

8 Clinicians

5 clinicians
3 radiologists, 1 neurosurgeon, 1 resident

3 clinicians
3 radiologists

Unaugmented Read

Washout Period

AI-Augmented Read

B. Original CTA Scan (Unaugmented Read)

C. Model Segmentation Overlay (AI-Augmented Read)

Unaugmented CTA Aneurysm Interpretation

AI-Augmented CTA Aneurysm Interpretation

Park & Chute & Rajpurkar et al, Jama Network Open, 2019
0.06 increase in sensitivity of detecting aneurysms.

Specificity difference is positive, but not significant.
Greatest improvement for clinician with lowest unaugmented score.

Smallest improvement for clinician with highest unaugmented score.
Deployment Challenges

- Algorithm Translation
- Clinical Impact

Engineering For Translation
Clinician Studies
Opportunities for AI in Medical Image Interpretation

Data

Deployment
Data Opportunities
● Fundamental Advances in Small Data and Transfer Learning

Deployment Opportunities
● ML Generalizability and Clinical Impact Studies
Data Opportunities

- Fundamental Advances in Small Data and Transfer Learning
Transfer learning on images effective

Pre-training on ImageNet

Fine-tuning on Medical Dataset
Transfer learning on video effective for small datasets
Class imbalance addressed to focus more on the hard-classified samples

\[ L(p(i|I), x) = \]
\[ - \frac{1}{|P|} \sum_{x_i=1} (1 - p(i|I))^\lambda \ln(p(i|I)) - \]
\[ - \frac{1}{|N|} \sum_{x_i=0} p(i|I)^\lambda \ln(1 - p(i|I)) \]

DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays, Chen et al., 2019
Idea comes from the Focal Loss Paper

\[
CE(p_t) = - \log(p_t)
\]

\[
FL(p_t) = -(1 - p_t)^\gamma \log(p_t)
\]

Focal Loss for Dense Object Detection, Lin et al., 2018
Transfer learning on CheXNet has little effect on performance

Transfusion: Understanding Transfer Learning with Applications to Medical Imaging, Raghu et al., 2019
Pre-Training is an effective transfer learning Strategy

<table>
<thead>
<tr>
<th>Strategy</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance on TB task with CheXpert pre-training</td>
<td>0.831 (0.751, 0.910)</td>
</tr>
<tr>
<td>Performance on TB task without CheXpert pre-training</td>
<td>0.714 (0.618, 0.810)</td>
</tr>
</tbody>
</table>

Rajpurkar & O’Connell & Schechter et al., *in submission*
Deployment Opportunities

- **ML Generalizability and Clinical Impact Studies**
Generalizability to Photo Deployment

patient00049/study1/view1

original x-ray

patient00410/study1/view1

original x-ray

original x-ray

original x-ray
Probability of Abnormality, ACL tear and MCL tear.

Bien & Rajpurkar et al, PLOS Medicine, 2018
Interactive AI Assistant

A Digital H&E slide

B

Upload to TumorAssistant

Extract Patch

C

Cancer Type A: 90%
Cancer Type B: 10%

Assisted Diagnosis

Rajpurkar & Uyumazturk & Kiani et al, 2019
Incorporation of Clinical Variables in Field Simulation

Rajpurkar & O’Connell & Schechter et al., *in submission*
Opportunities for AI + Medicine

Community
Open invite to the world to participate in competitions w/ hidden test set

2400 Users
80+ Teams Competing

1213 Users

3600 Users
70+ Teams Competing

Irvin & Rajpurkar et al. 2019, AAAI; Bien & Rajpurkar et al, PLOS Medicine; Rajpurkar & Irvin et al. 2018, MIDL
Weekly Newsletter on latest AI For Healthcare Research

1000+ Regular Weekly Readers

http://doctorpenguin.com
AI For Healthcare Bootcamp
AI For Healthcare Bootcamp

2-quarter program that provides Stanford students an opportunity to do cutting-edge research at the intersection of AI and healthcare.
AI For Healthcare Bootcamp

Training from PhD students and faculty in the medical school to work on high-impact research in small interdisciplinary teams.
https://stanfordmlgroup.github.io/programs/aihc-bootcamp/
or Google “AI For Healthcare Bootcamp Stanford”