DL for Healthcare

Pranav Rajpurkar
Advised by Andrew Ng
Stanford ML Group
Goals

Healthcare

What are high impact problems in healthcare that deep learning can solve?

Research

What does research in AI applications to medical imaging look like?

You

How can you get involved?
Goals

Healthcare

What are high impact problems in healthcare that deep learning can solve?
Questions we care about answering in healthcare

Descriptive
what happened?

What was a patient’s heart rate through their day?
Questions we care about answering in healthcare

Descriptive
what happened?

What was a patient’s heart rate through their day?

Diagnostic
why did it happen?

Why is this patient coughing for 2 weeks? Does their chest-xray show signs of pneumonia?

Why is this patient palpitating? Does their ECG show signs of atrial fibrillation?
Questions we care about answering in healthcare

**Descriptive**
what happened?

*What was a patient’s heart rate through their day?*

**Diagnostic**
why did it happen?

*Why is this patient coughing for 2 weeks? Does their chest-xray show signs of pneumonia?*

*Why is this patient palpitating? Does their ECG show signs of atrial fibrillation?*

**Predictive**
what will happen?

*Will this patient live for the next six months given their medical record?*

*Will this patient develop heart failure as a result of chemotherapy?*
### Questions we care about answering in healthcare

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Diagnostic</th>
<th>Predictive</th>
<th>Prescriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>what happened?</td>
<td>why did it happen?</td>
<td>what will happen?</td>
<td>what should we do?</td>
</tr>
<tr>
<td>What was a patient’s heart rate through their day?</td>
<td>Why is this patient coughing for 2 weeks? Does their chest-xray show signs of pneumonia?</td>
<td>Will this patient live for the next six months given their medical record?</td>
<td>Should this diabetic patient be treated with metformin or through lifestyle changes?</td>
</tr>
<tr>
<td></td>
<td>Why is this patient palpitating? Does their ECG show signs of atrial fibrillation?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Will this patient develop heart failure as a result of chemotherapy?</td>
<td></td>
</tr>
</tbody>
</table>
What was a patient’s heart rate through their day?

Why is this patient coughing for 2 weeks? Does their chest-xray show signs of pneumonia?

Why is this patient palpitating? Does their ECG show signs of atrial fibrillation?

Will this patient live for the next six months given their medical record?

Will this patient develop heart failure as a result of chemotherapy?

Should this diabetic patient be treated with metformin or through lifestyle changes?
Stanford ML Group

**Descriptive**
What happened?

- What was a patient’s heart rate through their day?

**Diagnostic**
Why did it happen?

- Why is this patient coughing for 2 weeks? Does their chest-xray show signs of pneumonia?
- Why is this patient palpitating? Does their ECG show signs of atrial fibrillation?

**Predictive**
What will happen?

- Will this patient live for the next six months given their medical record?
- Will this patient develop heart failure as a result of chemotherapy?

**Prescriptive**
What should we do?

- Should this diabetic patient be treated with metformin or through lifestyle changes?
What data can we use?

Medical Imaging
X-ray, CT, MRI, Ultrasound
What data can we use?

- **EHR**
  - Medications, lab tests, clinical notes

- **Medical Imaging**
  - X-ray, CT, MRI, Ultrasound
What data can we use?

- **EHR**
  - Medications, lab tests, clinical notes

- **Medical Imaging**
  - X-ray, CT, MRI, Ultrasound

- **Genomics**
  - DNA sequencing, RNA measurements
What data can we use?

- **Mobile**
  - Smartphones, wearables (ECG, PPG data)

- **EHR**
  - Medications, lab tests, clinical notes

- **Genomics**
  - DNA sequencing, RNA measurements

- **Medical Imaging**
  - X-ray, CT, MRI, Ultrasound
Where can decision making be assisted?

Clinician (radiologist, pathologist, physician)

Inside a hospital room

Medical Imaging
Where can decision making be assisted?

Inside a hospital room

- Medical Imaging

Across the hospital

- EHR

Clinician
- radiologist,
- pathologist,
- physician

Operations
Where can decision making be assisted?

- Inside a hospital room: Medical Imaging
- Across the hospital: EHR
- Beyond the hospital: Mobile

Clinician (radiologist, pathologist, physician)

Operations

Patients
How will decision making be affected?

Inform, Augment, or Replace

Clinician (radiologist, pathologist, physician)

Operations

Inside a hospital room

Medical Imaging

Across the hospital

EHR

Beyond the hospital

Mobile

Patients
Goals

Research

What does research in AI applications to medical imaging look like?
Deep Learning for Medical Imaging

Large Datasets → Deep Convolutional Neural Networks → Expert Level Performance
What’s tough?

- Large Datasets
  - Good labels?
  - How to acquire?
- Deep CNNs
  - ML process of developing networks
- Expert Level
  - Performance
  - Evaluating the importance
What's been done?

Gulshan et al., 2016

Ophthalmology
Detection of Diabetic Retinopathy

Dermatology
Detection of Melanomas

Esteva et al., 2017
What’s been done?

- Dermatology: Detection of Melanomas
- Cardiology: Detection of Arrhythmias
- Radiology: Detection of Pneumonia
- Ophthalmology: Detection of Diabetic Retinopathy

- Gulshan et al., 2016
- Rajpurkar et al., 2017
- Esteva et al., 2017
- Rajpurkar et al., 2017
CheXNet
Radiologist-Level Pneumonia Detection on Chest X-Rays

Pranav Rajpurkar*, Jeremy Irvin*, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng
Pneumonia detection is important

Infection that inflames the air sacs in lungs.

1 million hospitalizations and 50,000 deaths per year in the US alone.

Symptoms: cough with phlegm, fever, chills, trouble breathing. Like people with colds or the flu, but lasts longer.
To Diagnose

Diagnosis starts with symptoms and a stethoscope.

If signs of pneumonia, then take an x-ray.
Chest X-ray exam

Fast and painless imaging test using x-rays.

Usually two views, one from straight on and one from the side of chest.

2 billion chest x-ray procedures per year.
Chest X-ray image

Ribs and spine will absorb much of the radiation and appear white or light gray on the image.

Lung tissue absorbs little radiation and will appear dark on the image.

Air appears black.
Detecting Abnormalities

Abnormalities present mostly as areas of increased density (opacity).
X-ray findings of pneumonia

Most commonly manifests as consolidation ("fluffy cloud").
X-ray findings of pneumonia

Most commonly manifests as consolidation ("fluffy cloud").

Lobar pneumonia: entire lobe consolidated.
X-ray findings of pneumonia

Most commonly manifests as consolidation ("fluffy cloud").

Lobar pneumonia: entire lobe consolidated.
Detecting Pneumonia

Pneumonia occurs when alveoli fill up with pus.
Confusing Pneumonia

Appearance of pneumonia in X-ray images is often vague, and can mimic other abnormalities.

If not pus filling up alveoli, but:
- Cells (cancer)
- Blood (pulmonary hemorrhage)
Contributions

1. Radiologist-level pneumonia detection from Chest X-rays.

2. State of the art results on all 14 thoracic pathologies in the largest public x-ray dataset.
Setup

- Input is a frontal view chest X-ray image
- Output is a binary label $t \in \{0, 1\}$ indicating the absence or presence of pneumonia
Network Architecture

- 2D CNN over 224 x 224 images
- Pretrained on ImageNet
- 121 layer DenseNet Architecture
DenseNets

- Connect every layer to every other layer in feed forward fashion

Densely Connected Convolutional Networks Huang & Liu et al. (2016)
DenseNets

- Beats previous state of the art (ResNet) and have:
  - lower error
  - fewer parameters

Densely Connected Convolutional Networks Huang & Liu et al. (2016)
Dataset
Building off of public x-ray scans
Dataset

- 112,120 frontal-view X-ray images of 30,805 unique patients
- Largest public dataset (released sep 2017)

ChestX-ray14
Wang et al. (2017)
Dataset - Train Set

- Each x-ray annotated with up to 14 different thoracic pathology labels
- Annotation by NLP on radiology reports

ChestX-ray14
Wang et al. (2017)
Dataset - Test Set

- We collected a test set of 420 frontal chest X-rays.
- 4 Stanford radiologists independently annotated
Lots of data & deep network

How close to experts can we get?
Evaluation -- Metrics

precision = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false positives}}

recall = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}

**Goal:** maximize both precision and recall

\[ F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]
Evaluation -- Assessing Radiologists

For each radiologist, we calculate their F1-score using each of the other three radiologists, and CheXNet, as ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing Radiologists

For each radiologist, we calculate their F1-score using each of the other three radiologists, and CheXNet, as ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing Radiologists

For each radiologist, we calculate their F1-score using each of the other three radiologists, and CheXNet, as ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing Radiologists

For each radiologist, we calculate their F1-score using each of the other three radiologists, and CheXNet, as ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing CheXNet

For our model, we calculate F1-score using the each of the four radiologists as the ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing CheXNet

For our model, we calculate F1-score using the each of the four radiologists as the ground truth.

Repeat for test set (420 images)
For our model, we calculate F1-score using the each of the four radiologists as the ground truth.

Repeat for test set (420 images)
Evaluation -- Assessing CheXNet

For our model, we calculate F1-score using the each of the four radiologists as the ground truth.

Repeat for test set (420 images)
## Evaluation -- Results

<table>
<thead>
<tr>
<th>Radiologist</th>
<th>F1 Score (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiologist 1</td>
<td>0.383 (0.309, 0.453)</td>
</tr>
<tr>
<td>Radiologist 2</td>
<td>0.356 (0.282, 0.428)</td>
</tr>
<tr>
<td>Radiologist 3</td>
<td>0.365 (0.291, 0.435)</td>
</tr>
<tr>
<td>Radiologist 4</td>
<td>0.442 (0.390, 0.492)</td>
</tr>
<tr>
<td>Radiologist Avg.</td>
<td>0.387 (0.330, 0.442)</td>
</tr>
<tr>
<td>CheXNet</td>
<td>0.435 (0.387, 0.481)</td>
</tr>
</tbody>
</table>
We identify two limitations with our comparison to radiologists:

1. No access to patient history or prior examinations.
2. Only frontal radiographs presented, no lateral views.
### Evaluation -- Previous Benchmarks

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Wang et al. (2017)</th>
<th>Yao et al. (2017)</th>
<th>CheXNet (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelectasis</td>
<td>0.716</td>
<td>0.772</td>
<td>0.8094</td>
</tr>
<tr>
<td>Cardiomegaly</td>
<td>0.807</td>
<td>0.904</td>
<td>0.9248</td>
</tr>
<tr>
<td>Effusion</td>
<td>0.784</td>
<td>0.859</td>
<td>0.8638</td>
</tr>
<tr>
<td>Infiltration</td>
<td>0.609</td>
<td>0.695</td>
<td>0.7345</td>
</tr>
<tr>
<td>Mass</td>
<td>0.706</td>
<td>0.792</td>
<td>0.8676</td>
</tr>
<tr>
<td>Nodule</td>
<td>0.671</td>
<td>0.717</td>
<td>0.7802</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>0.633</td>
<td>0.713</td>
<td>0.7680</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>0.806</td>
<td>0.841</td>
<td>0.8887</td>
</tr>
<tr>
<td>Consolidation</td>
<td>0.708</td>
<td>0.788</td>
<td>0.7901</td>
</tr>
<tr>
<td>Edema</td>
<td>0.835</td>
<td>0.882</td>
<td>0.8878</td>
</tr>
<tr>
<td>Emphysema</td>
<td>0.815</td>
<td>0.829</td>
<td>0.9371</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>0.769</td>
<td>0.767</td>
<td>0.8047</td>
</tr>
<tr>
<td>Pleural Thickening</td>
<td>0.708</td>
<td>0.765</td>
<td>0.8062</td>
</tr>
<tr>
<td>Hernia</td>
<td>0.767</td>
<td>0.914</td>
<td>0.9164</td>
</tr>
</tbody>
</table>

Evaluated by AUROC in the binary classification tasks for each of the 14 pathologies.
XRay4All
With Michael Bereket, Thao Nguyen, and Henrik Marklund

XRay4All: Making X-Ray Diagnoses Quick and Accessible through AI

Diagnosis
We diagnose 14 pathologies from any Chest-Xray images.

World-Class Algorithm
Re-implements algorithm that achieved radiologist-level performance in diagnosing pneumonia from X-rays.

Global Impact
There is a shortage of skilled radiologists: 2/3 of the world doesn't have access to their radiology needs. This is free of cost and takes 0.1 seconds.
Rivaling clinical experts!
How do we interpret the algorithm?
Model Interpretation

Can you trust your model?

What parts of an image are most important for diagnosis?
Class Activation Maps

Learning deep features for discriminative localization Zhou et al. (2016)
Pneumonia

Multifocal community acquired pneumonia

Left lower and right upper lobes
Pneumothorax

Right-sided pneumothorax
Nodules

Left lower lung nodule

90% of mistakes in lung cancer diagnosis occurs on chest radiographs
AI for pneumonia detection from chest x-rays

Can it make an impact?
Future of diagnostic access

1. Improve healthcare delivery.

   CheXNet can help radiologists prioritize workflow and make better diagnoses.

2. Increase access to medical imaging expertise globally.

   $\frac{2}{3}$ of the global population lack access to radiology diagnostics.
Goals

You

How can you get involved?
AI for Healthcare Bootcamp with Andrew Ng

For ML students intending to get involved in research

2-quarter bootcamp covers a large breadth of topics at the intersection of artificial intelligence and healthcare. Students take a dive into cutting-edge research in AI for healthcare.
Next bootcamp in Fall. Applications open today!

https://stanfordmlgroup.github.io/programs/aihc-bootcamp-fall2018/