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?

Diagnostic
why did it happen?

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?

Questions we care about answering in healthcare

Descriptive what happened?	Diagnostic why did it happen?	Predictive what will happen?
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?

Questions we care about answering in healthcare

Descriptive what happened?	Diagnostic why did it happen?	Predictive what will happen?	Prescriptive what should we do?
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?

Yesterday

Descriptive
what happened?

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

Today

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?

Today

Predictive
what will happen?

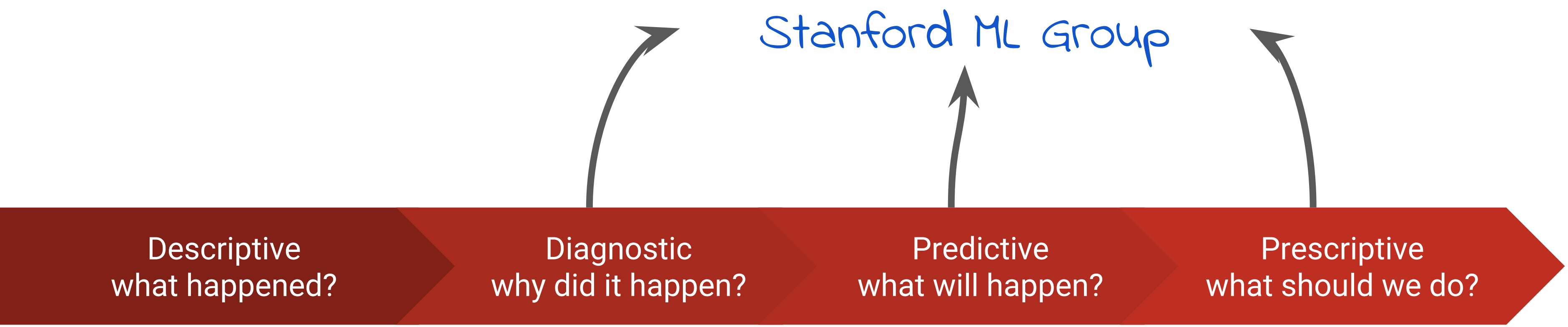
Will this patient live for the next six months given their medical record?

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

Tomorrow

Prescriptive
what should we do?

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



Descriptive

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

Diagnostic

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

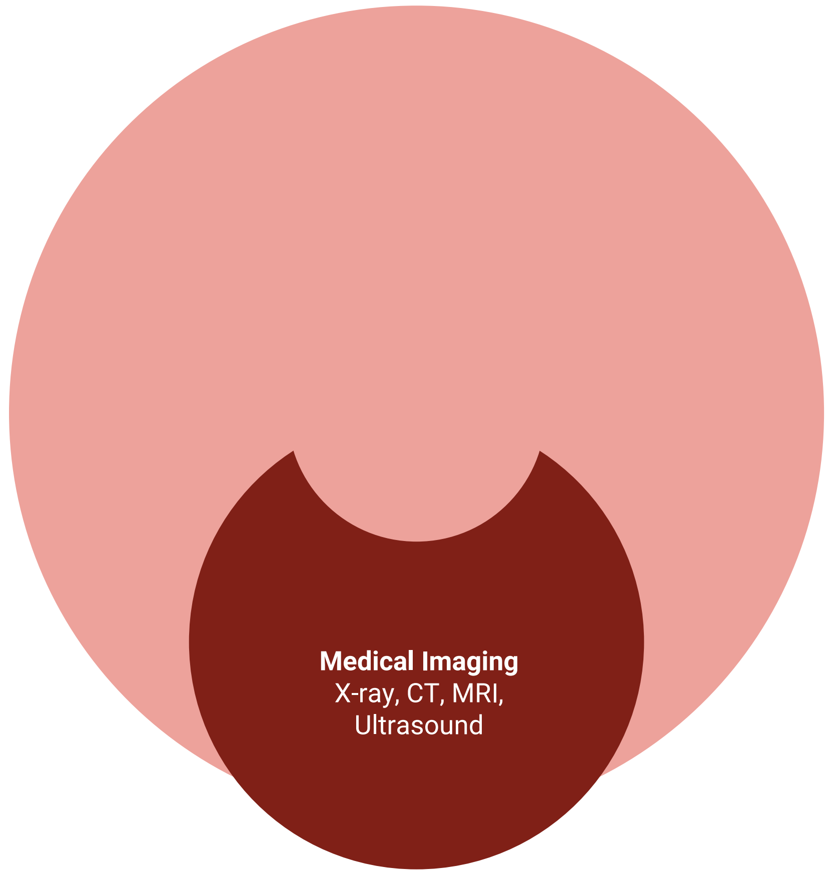
Will this patient live for the next six months given their medical record?

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

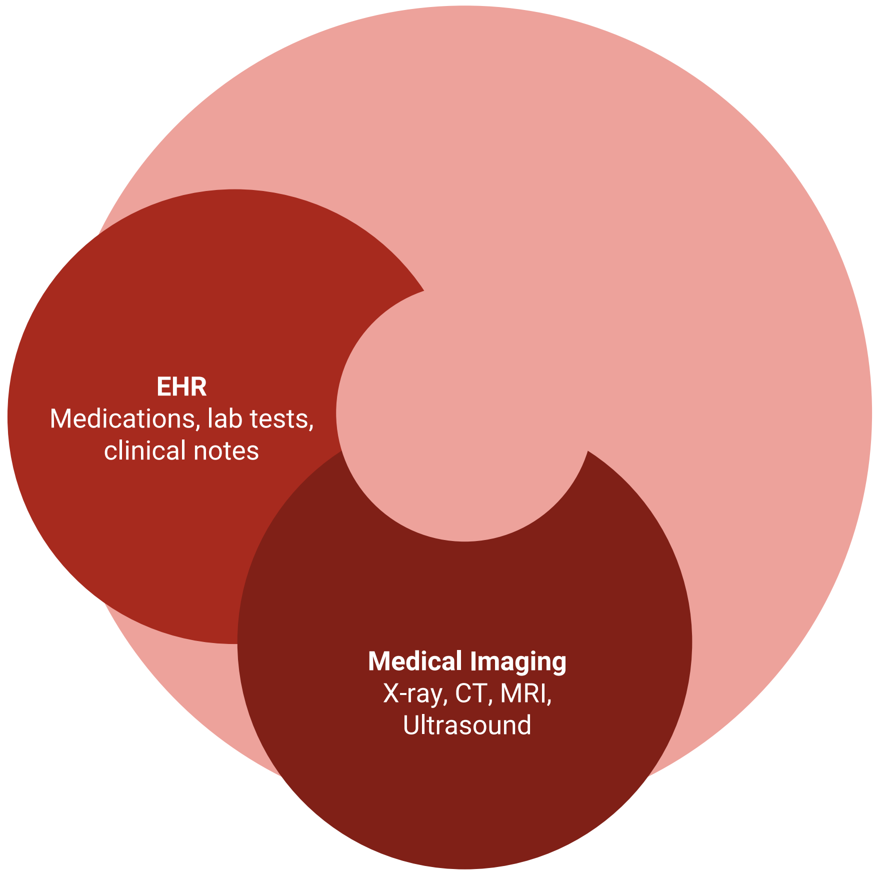
Prescriptive

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

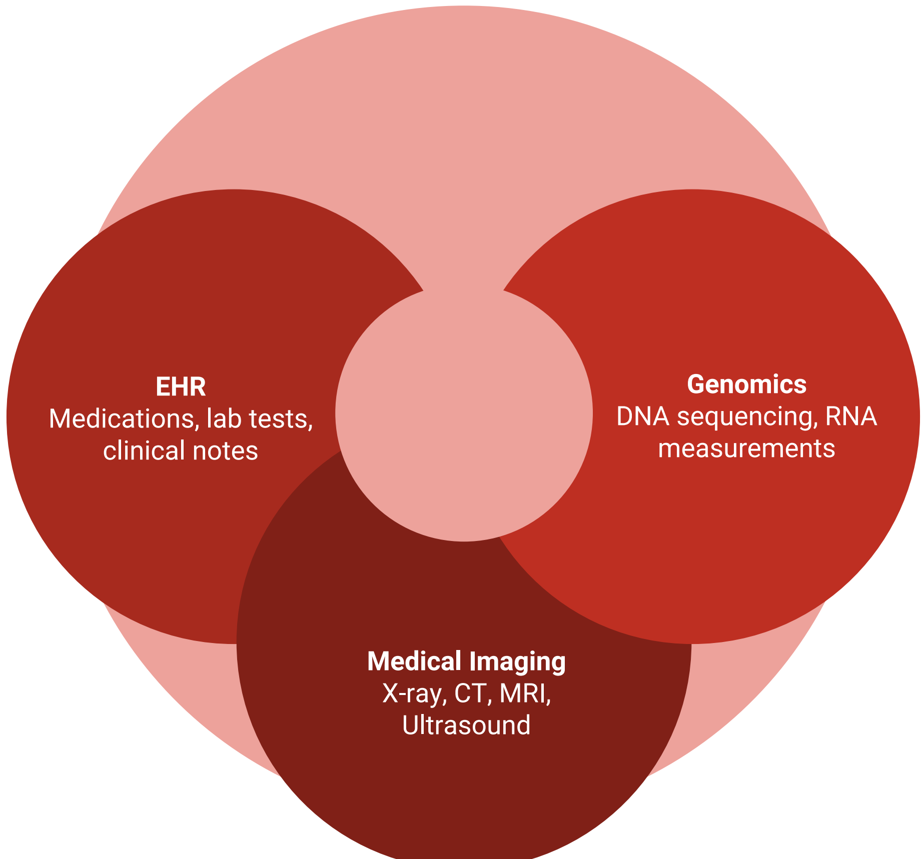
What data can we use?



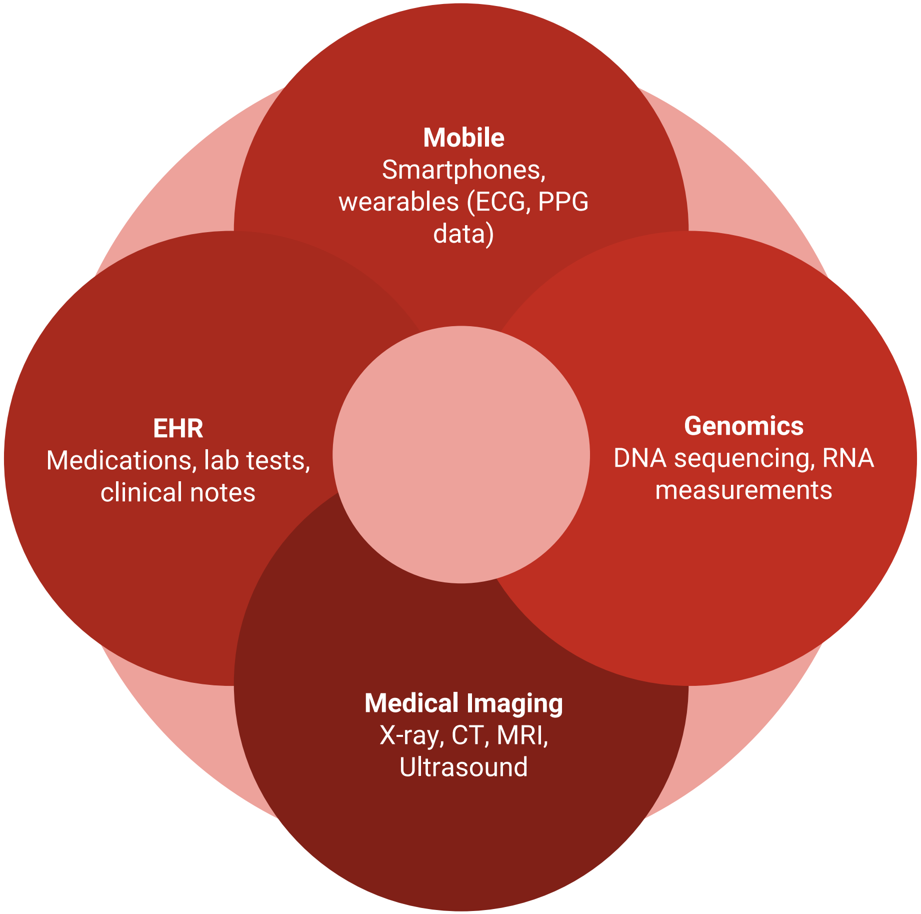
What data can we use?



What data can we use?

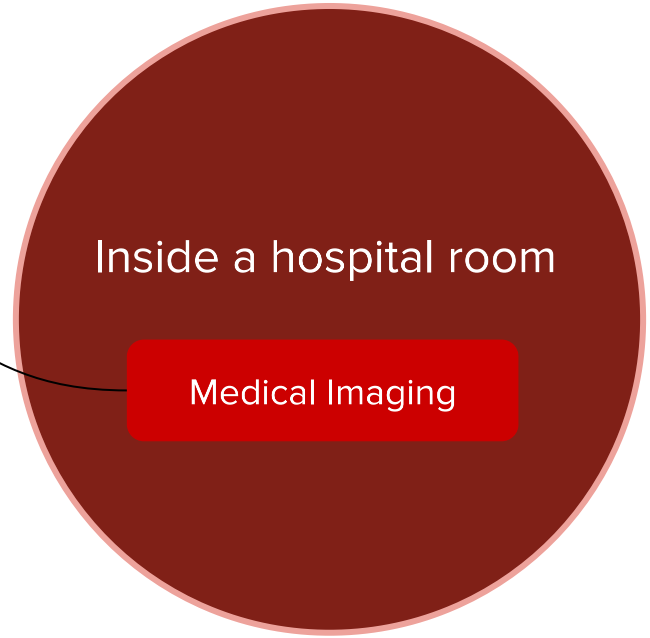


What data can we use?

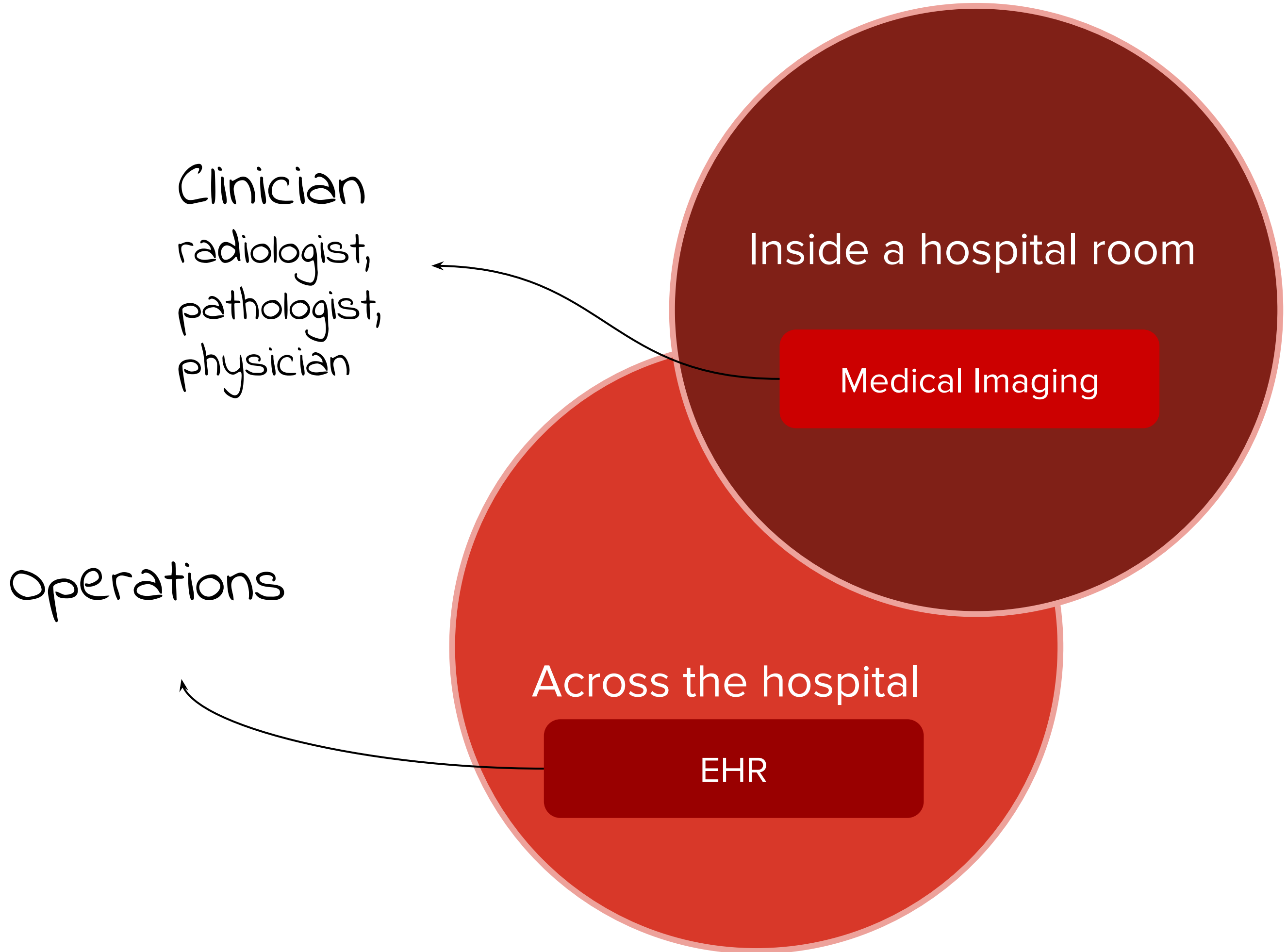


Where can decision making be assisted?

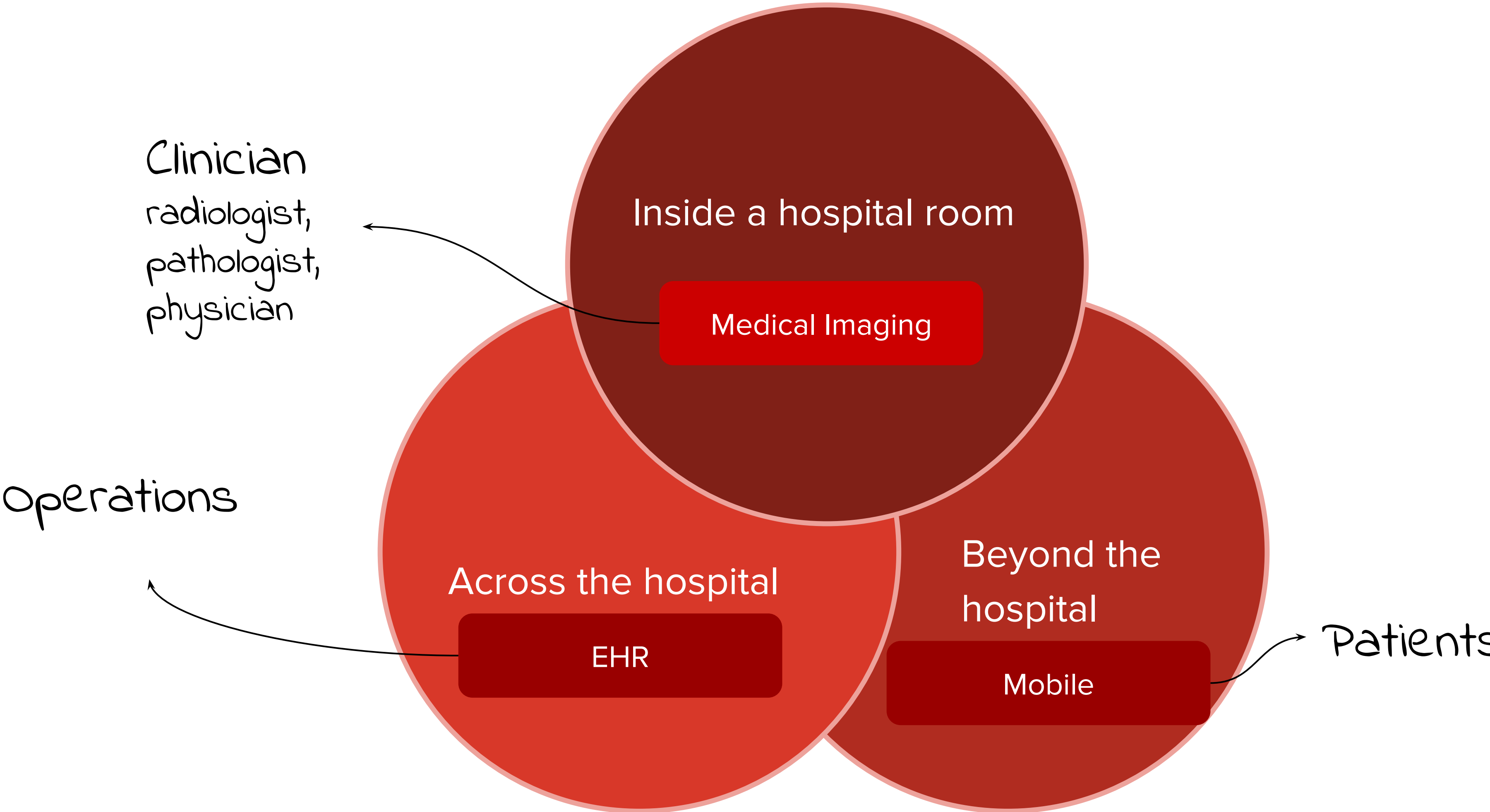
Clinician
radiologist,
pathologist,
physician



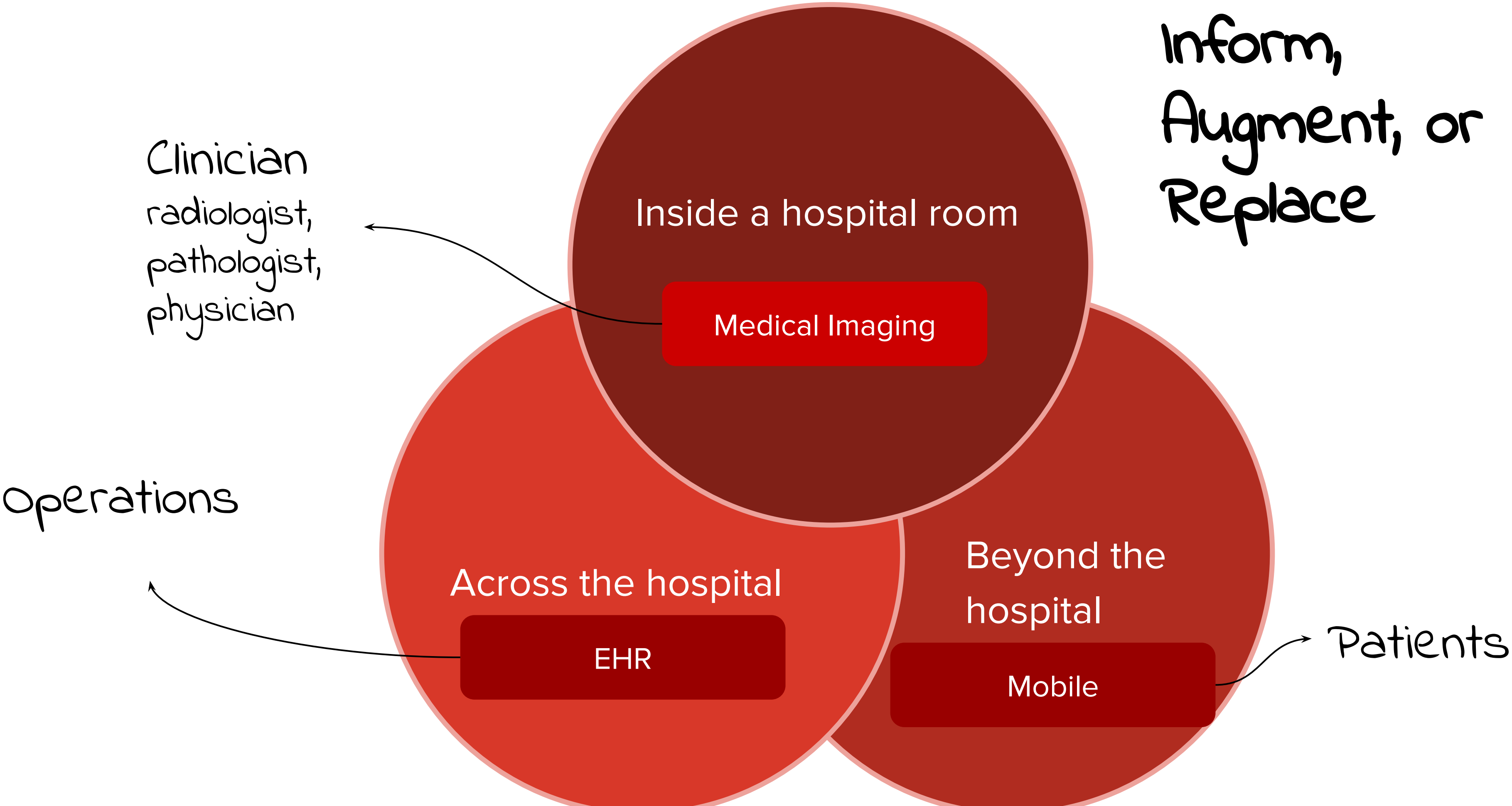
Where can decision making be assisted?



Where can decision making be assisted?



How will decision making be affected?

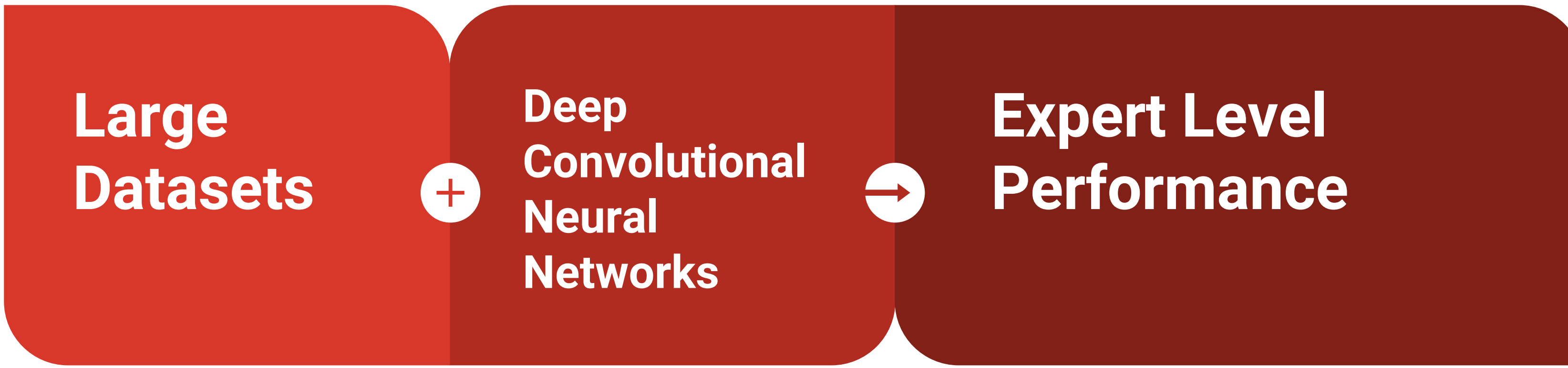


Goals

Research

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

Deep Learning for Medical Imaging



What's tough?

Large Datasets
Good labels?
How to acquire?

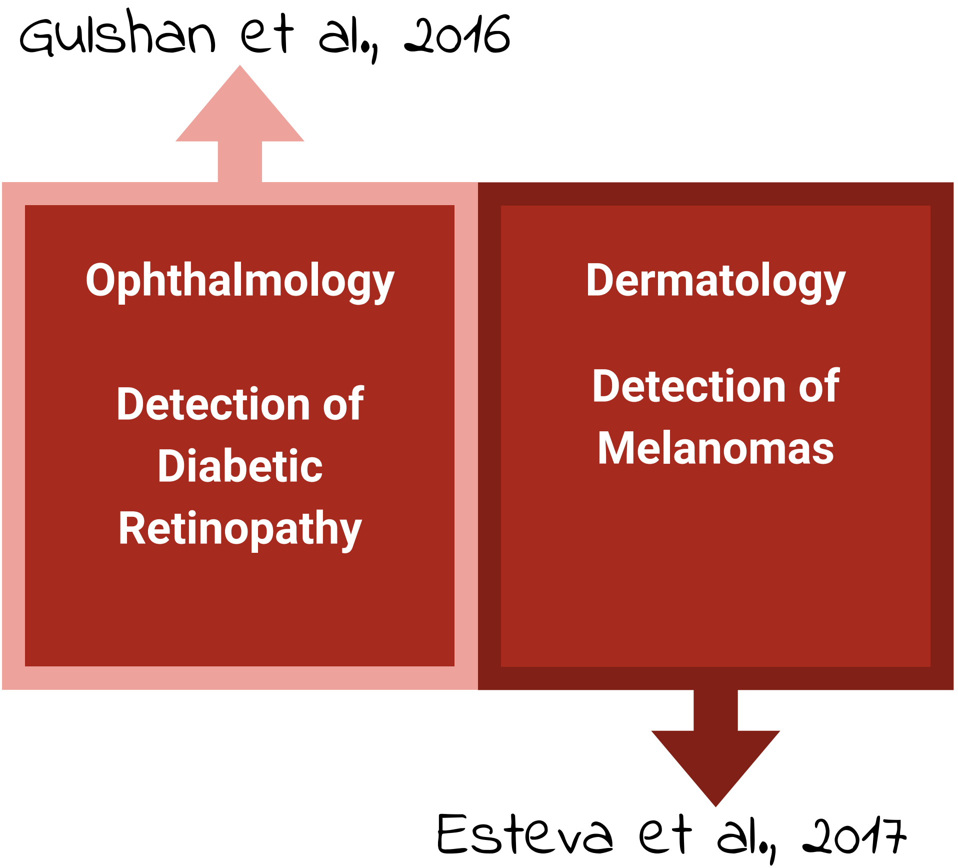
+

Deep CNNs
ML process of
developing networks

→

Expert Level Performance
Evaluating the importance

What's been done?



What's been done?

Gulshan et al., 2016

Ophthalmology

Detection of Diabetic Retinopathy

Esteva et al., 2017

Dermatology

Detection of Melanomas

Rajpurkar et al., 2017

Cardiology

Detection of Arrhythmias

Rajpurkar et al., 2017

Radiology

Detection of Pneumonia

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.

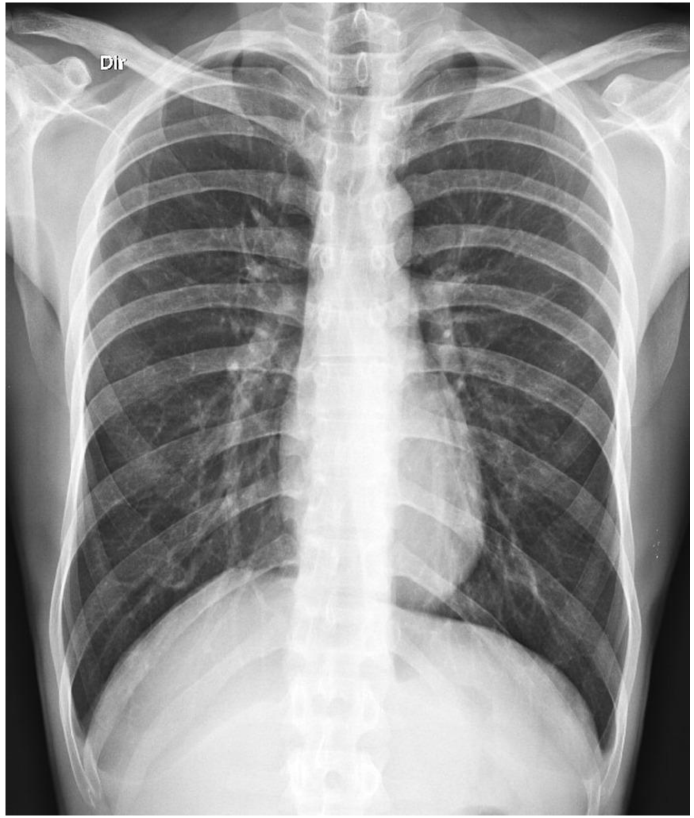
2 **billion** chest x-ray procedures per year.

Chest X-ray image

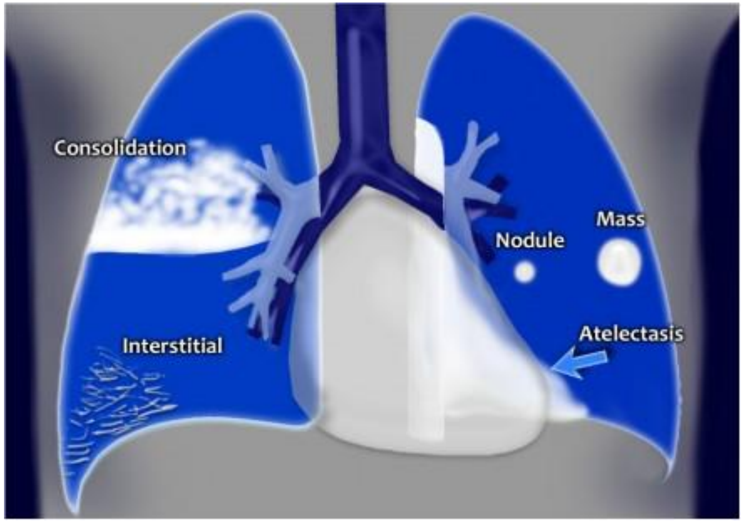
Ribs and spine will absorb much of the radiation and appear white or light gray on the image.

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

Air appears black.



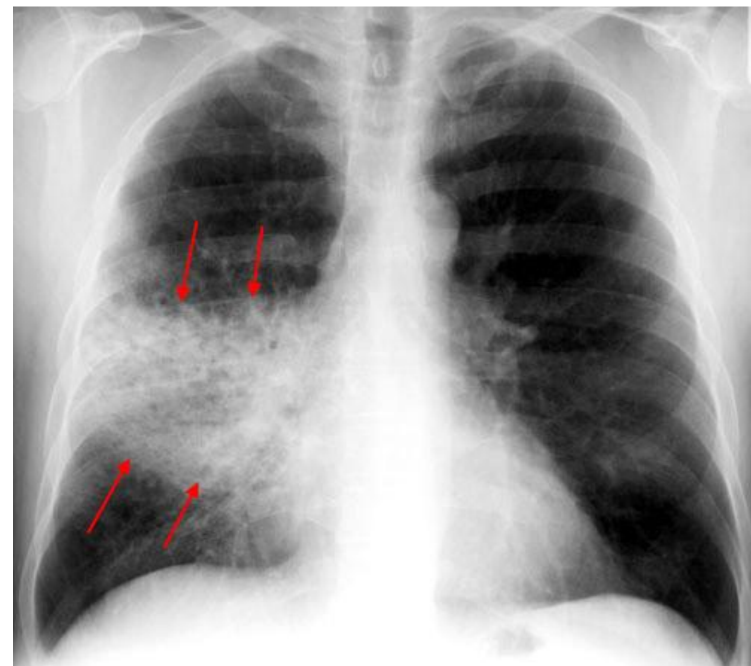
Detecting Abnormalities



Abnormalities present mostly as areas of **increased density** (opacity).

X-ray findings of pneumonia

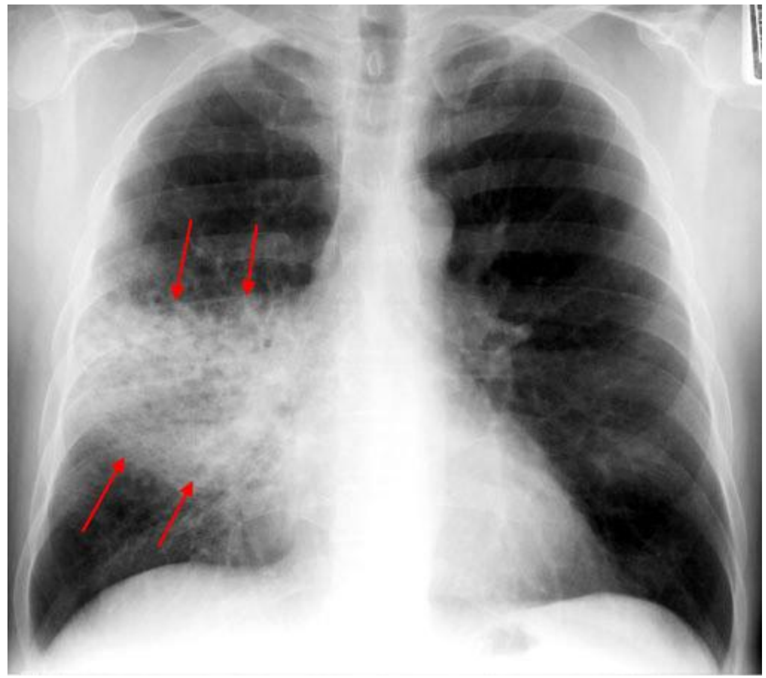
Most commonly manifests as consolidation ("fluffy cloud").



X-ray findings of pneumonia

Most commonly manifests as consolidation (“fluffy cloud”).

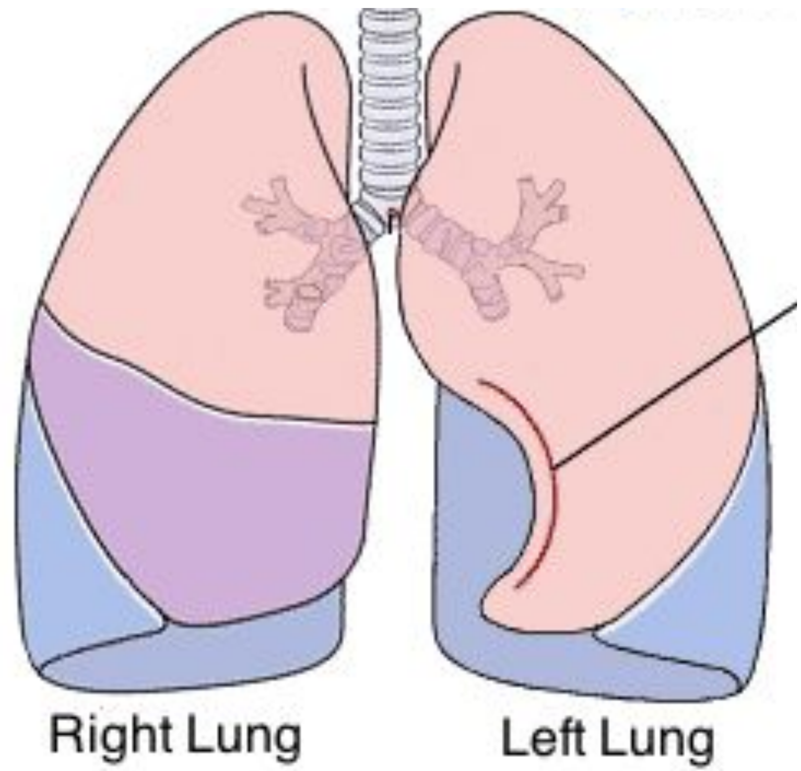
Lobar pneumonia: entire lobe consolidated.



X-ray findings of pneumonia

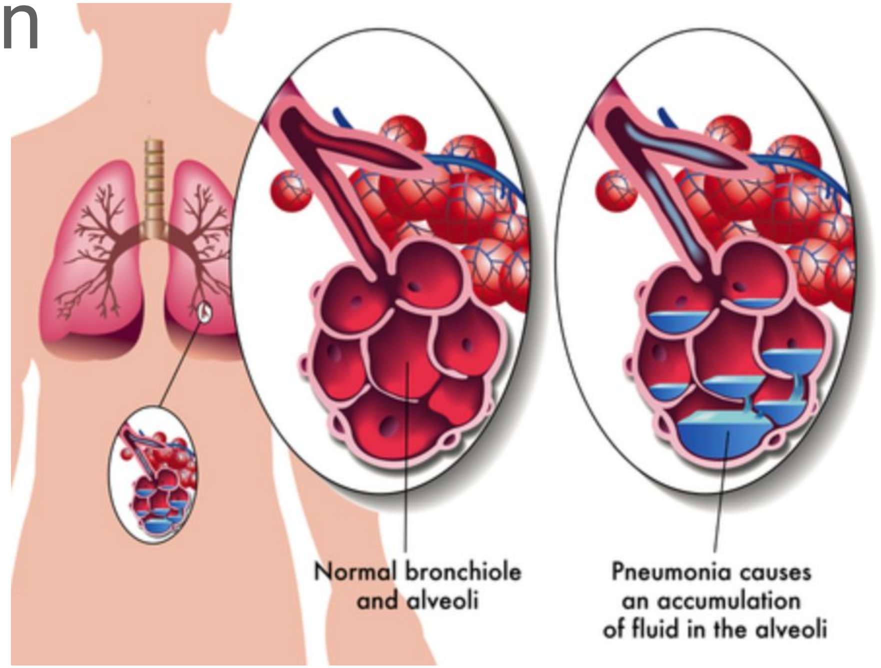
Most commonly manifests as consolidation (“fluffy cloud”).

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)

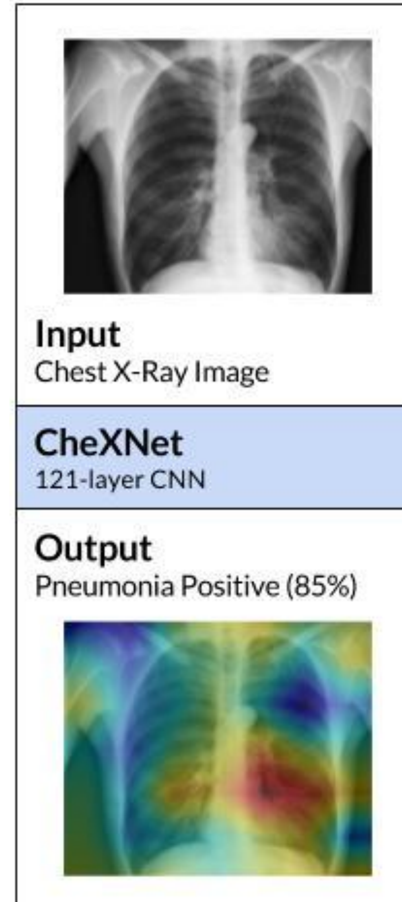
Contributions

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.

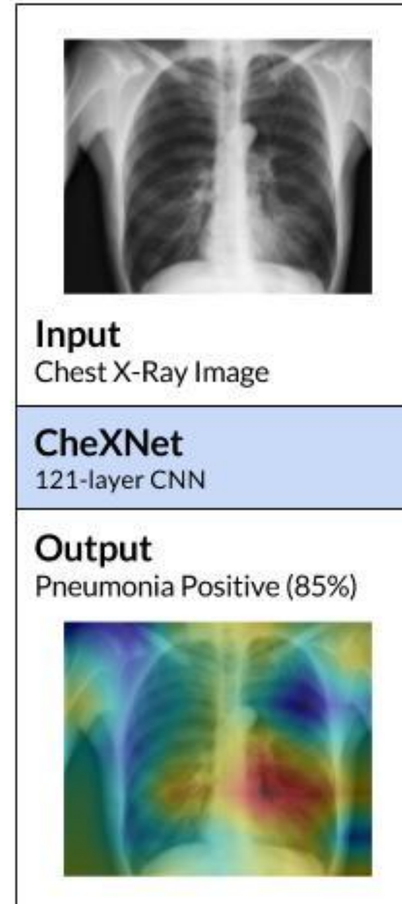
Setup

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



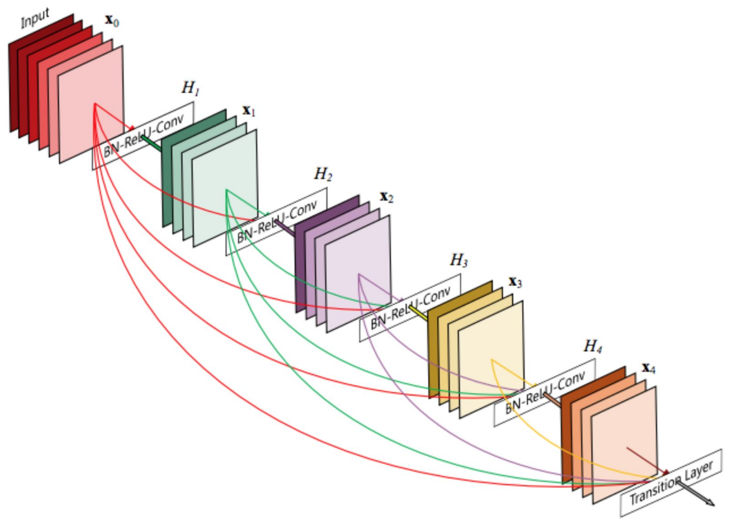
Network Architecture

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



DenseNets

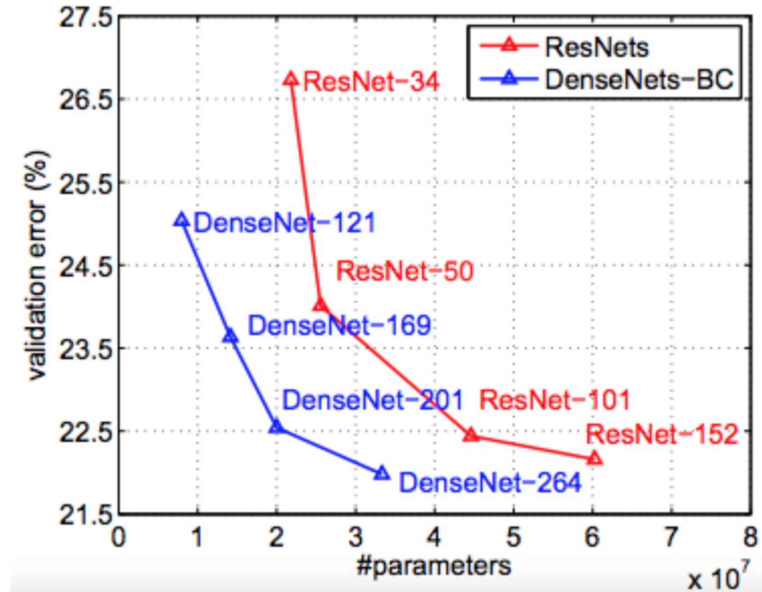
- Connect every layer to every other layer in feed forward fashion



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

DenseNets

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



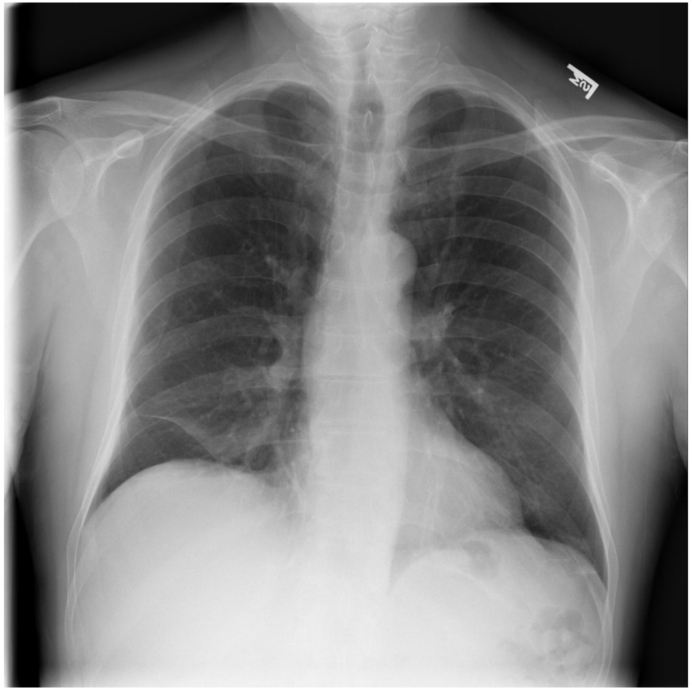
Densely Connected Convolutional Networks **Huang & Liu et al. (2016)**

Dataset

Building off of public x-ray scans

Dataset

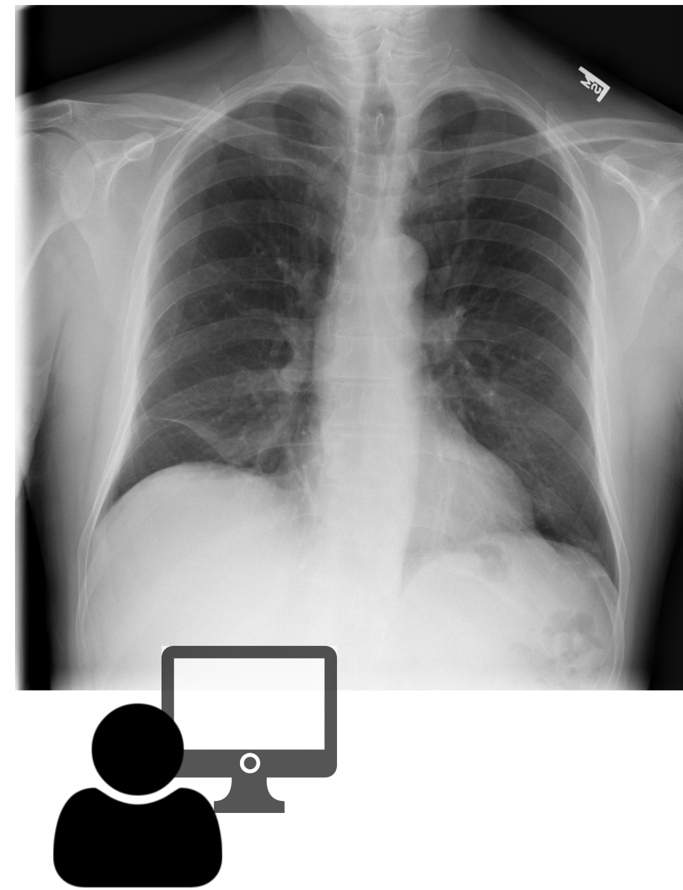
- 112,120 frontal-view X-ray images of 30,805 unique patients
- Largest public dataset (released sep 2017)



ChestX-ray14
Wang et al. (2017)

Dataset - Train Set

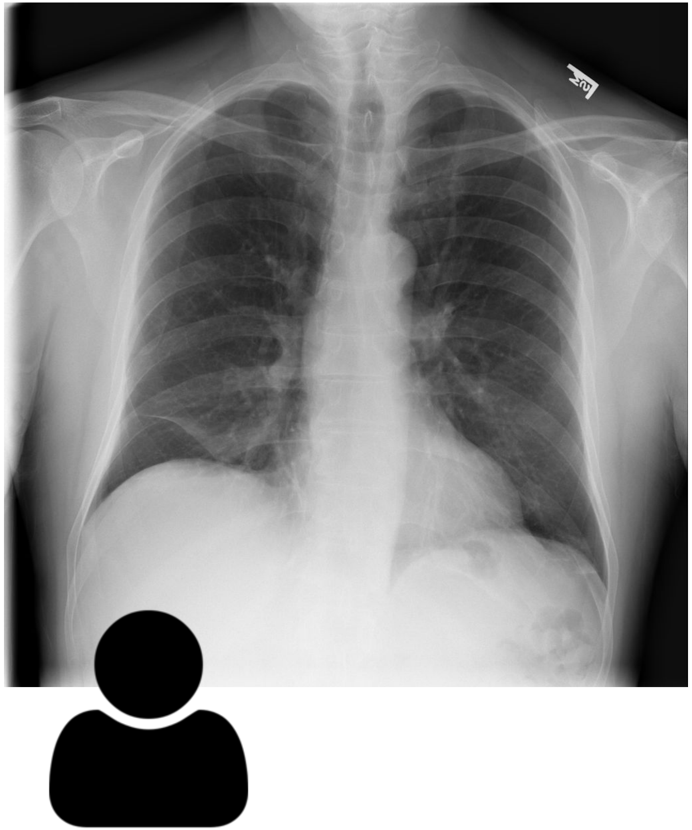
- Each x-ray annotated with up to 14 different thoracic pathology labels
- Annotation by NLP on radiology reports



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

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

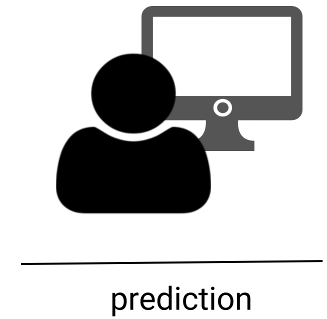
$$\text{recall} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

Goal: maximize both precision and recall

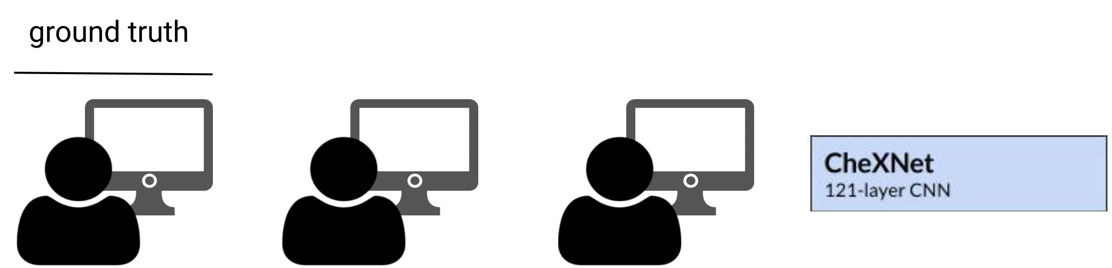
$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Evaluation -- Assessing Radiologists

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

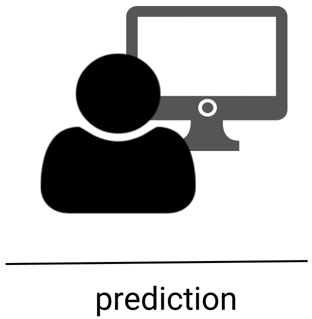


Repeat for test set
(420 images)

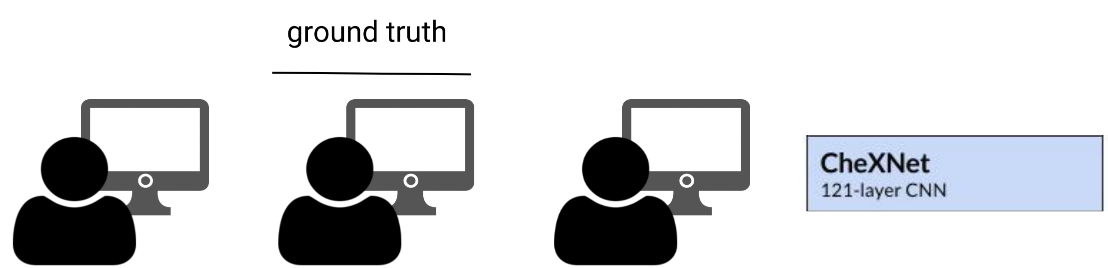


Evaluation -- Assessing Radiologists

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

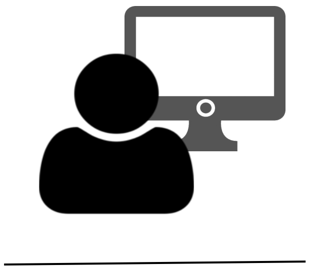


Repeat for test set **(420 images)**

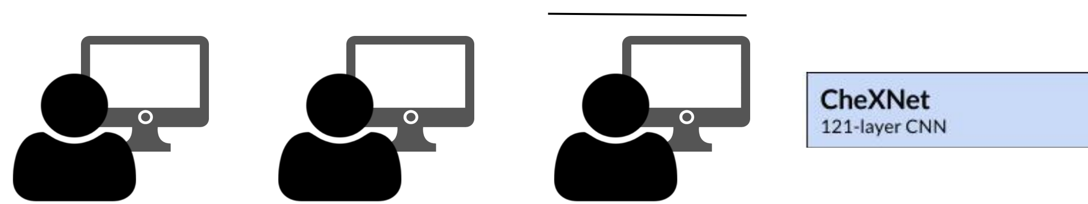


Evaluation -- Assessing Radiologists

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

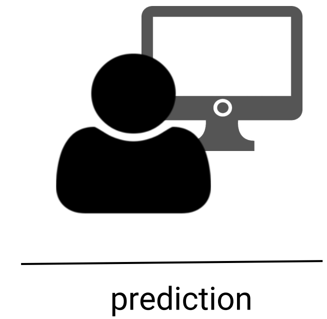


Repeat for test set
(420 images)

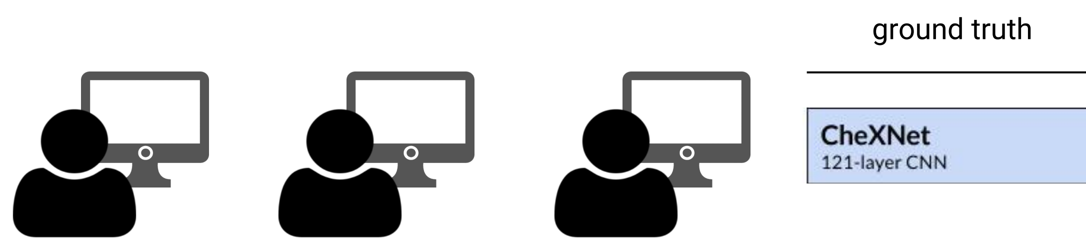


Evaluation -- Assessing Radiologists

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

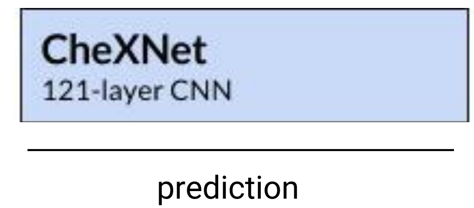


Repeat for test set
(420 images)

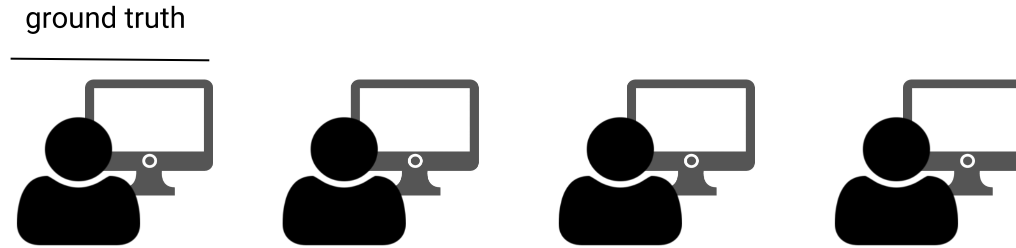


Evaluation -- Assessing CheXNet

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



Repeat for test set
(420 images)



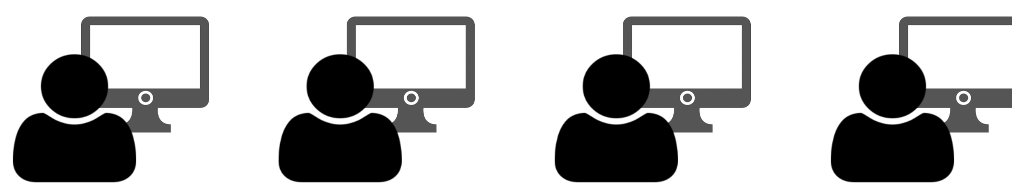
Evaluation -- Assessing CheXNet

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

CheXNet
121-layer CNN

prediction

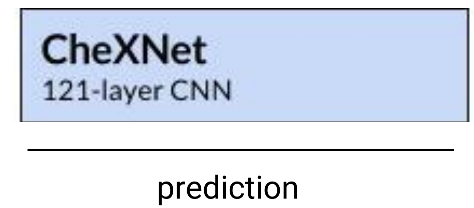
ground truth



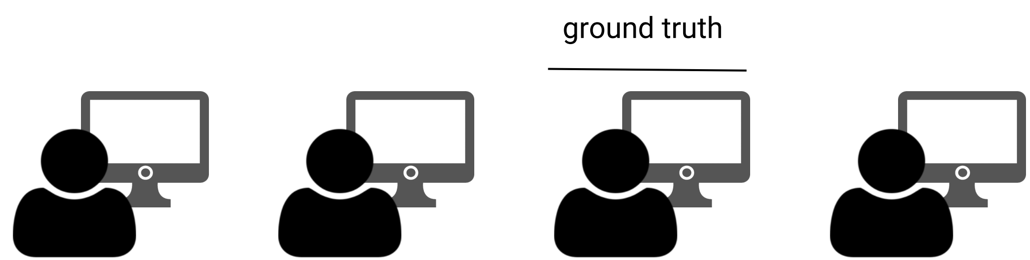
Repeat for test set
(420 images)

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

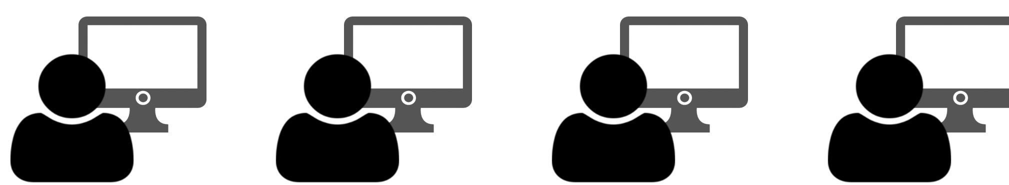
For our model, we calculate F1-score using the each of the four radiologists as the ground truth.

CheXNet
121-layer CNN

prediction

Repeat for test set
(420 images)

ground truth



Evaluation -- Results

	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

Evaluation -- Limitations

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

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

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

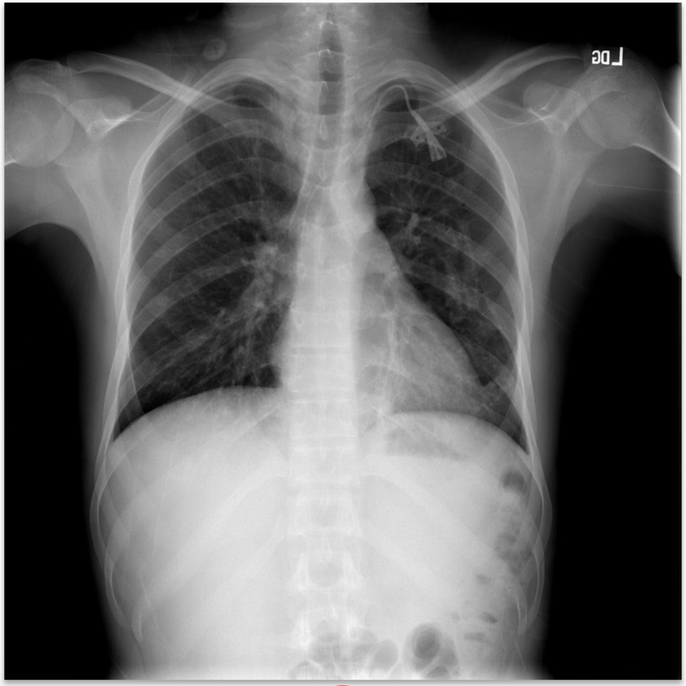
We diagnose 14 pathologies from any Chest-Xray images.

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!

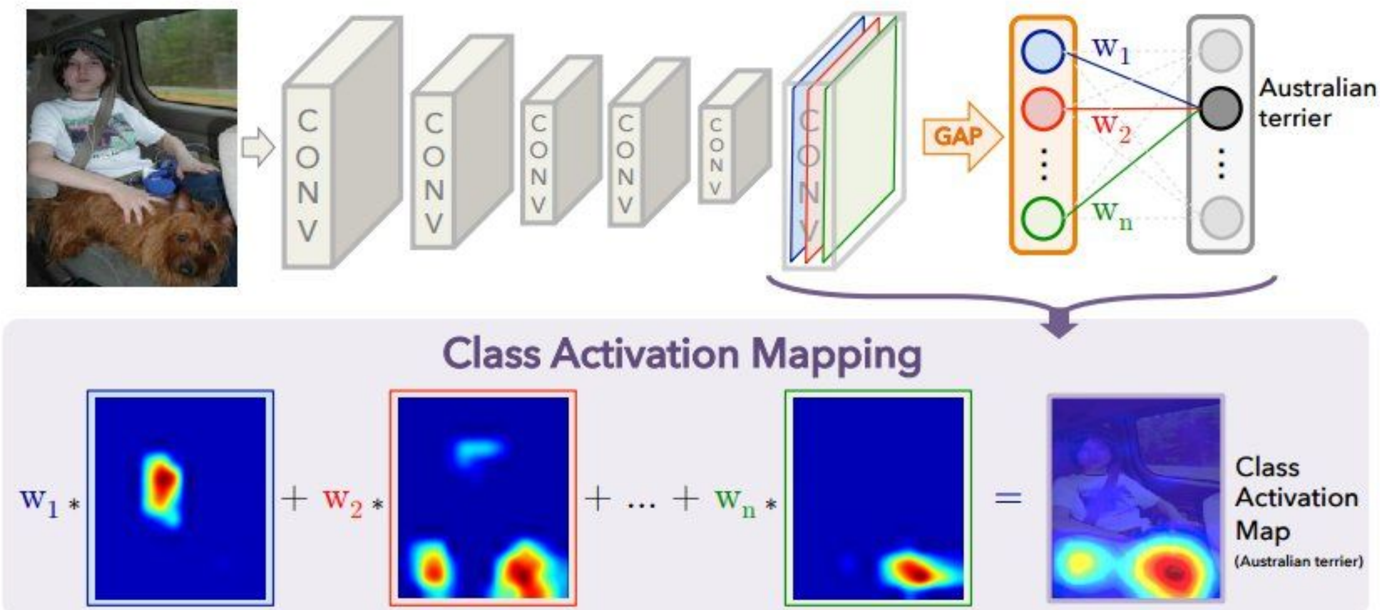
How do we interpret the algorithm?

Model Interpretation

Can you trust your model?

What parts of an image are most important for diagnosis?

Class Activation Maps

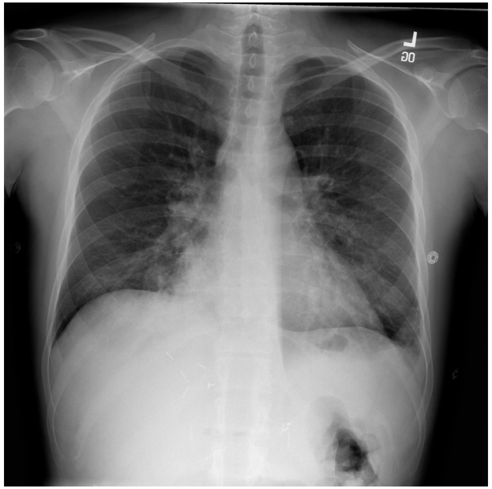


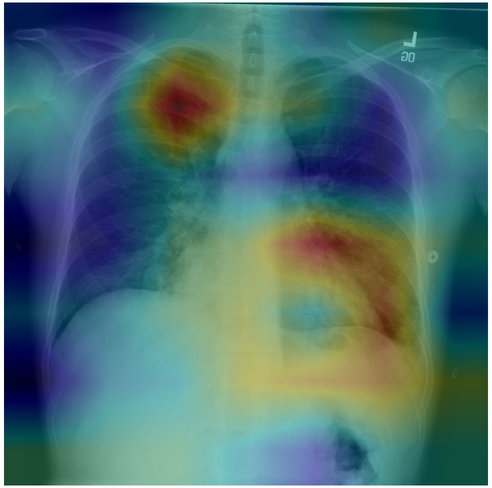
Learning deep features for discriminative localization **Zhou et al. (2016)**

Pneumonia

Multifocal community acquired pneumonia

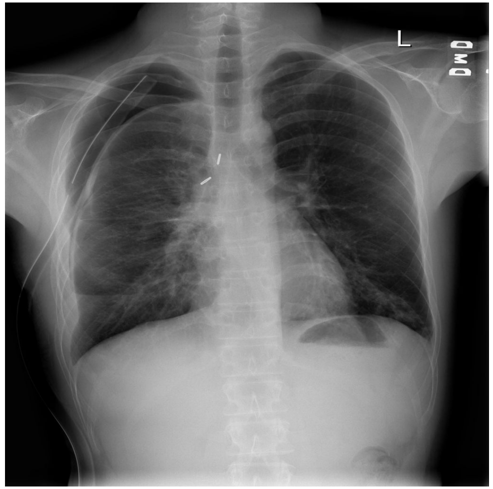
Left lower and right upper lobes

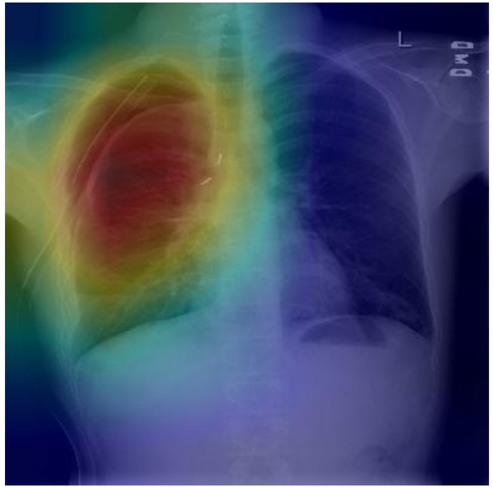




Pneumothorax

Right-sided pneumothorax

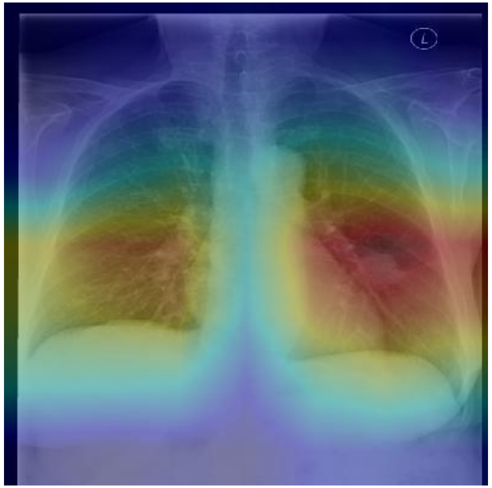




Nodules

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.

   ⅔ of the global population lack access to radiology diagnostics.

Goals

You

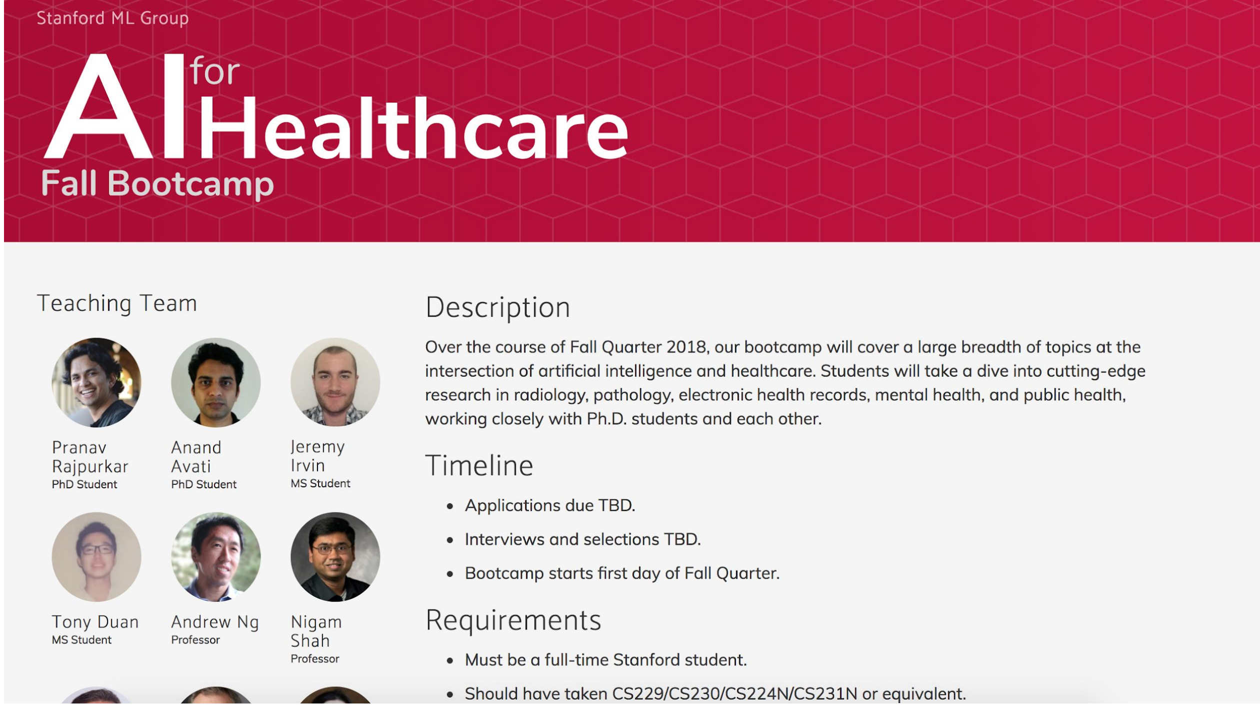
How can you get involved?

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/