NIH CHEST X-RAY

Our Puzzle:

Does subsetting data improve the performance of ML models?

Group members:
Lyndon Lee
Ziyi Yan
Michael McCarty

INTRODUCTION

Motivations

A number of **medical conditions** may be diagnosed with a **chest x ray**, and highly trained radiologists analyze the images for **indications**.

While some conditions are fairly easy to diagnose, others are very **difficult**, with even experts disagreeing on the diagnosis.

With the success of AI in other image recognition areas using **CNNs**, this is an area ripe for **AI assistance**. Even if the AI works comparatively similar to a human, can be useful for triage.

Data Provenance

- The National Institute of Health released a large dataset of **over 112,000 labelled chest x ray images**
- All personally identifiable information was removed, but **age and gender** are included.
- labelled with **14 common chest conditions**, many images with **more than one**.
- Labels were determined by using **NLP** on the associated radiological reports (expected to have **accuracy >90%**)

ISSUES WITH X RAY IMAGES

- X rays can be taken from either **Anterior-to-Posterior (AP), or Posterior-to-Anterior (PA)**.

- Generally, **PA images are preferred**, but for a sick patient that is unable to stand, AP images are taken.
    - For example, on AP images, the heart looks artificially larger than in reality, which makes cardiomegaly (enlarged heart) difficult to diagnose.

- Females can present issues as the images can show **breast tissue** which can make images more difficult to read.

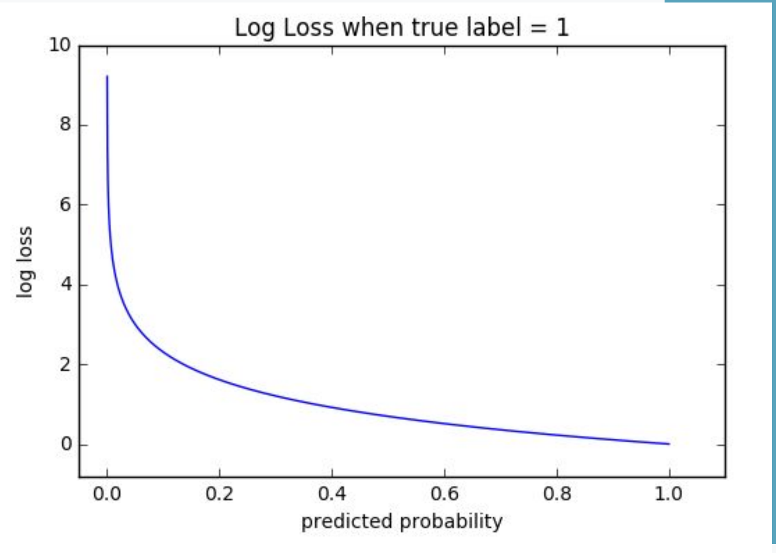
METHODOLOGY

Subsetting Data

- Data subset into **2x2 = 4 groups** based on **sex** and image **view orientation**:
- We then compare **ROC values** to track if we can find significant improvements

Male & AP	Male & PA
Female & AP	Female & PA

Loss Function:
Binary Cross Entropy



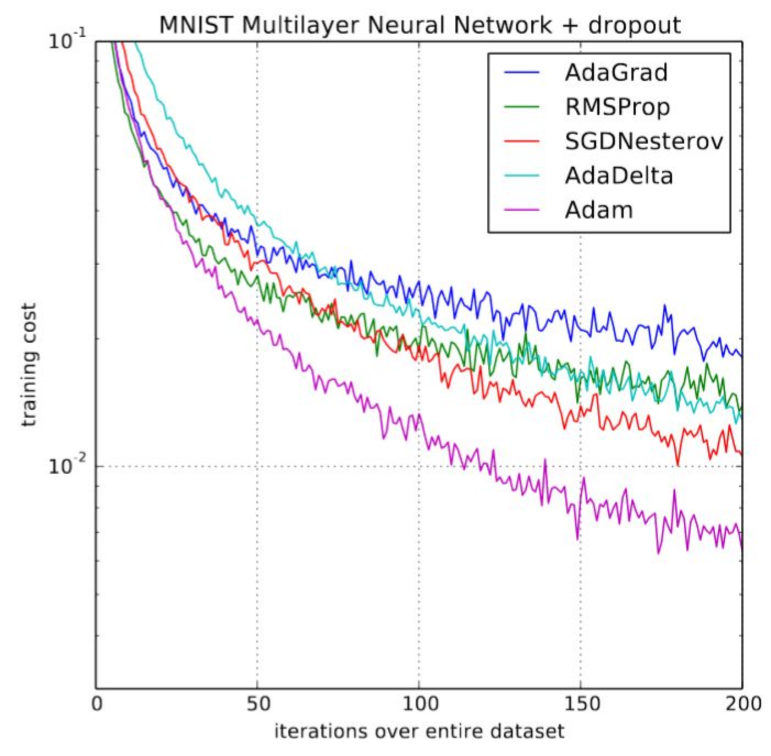
$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Negative log probability: closer the probability is to 1 (more likely it is the image), the smaller the penalty

Did **not** use **Categorical Cross-Entropy** because each image can be labelled with more than one disease: binary choice for assessing each image for 1 disease

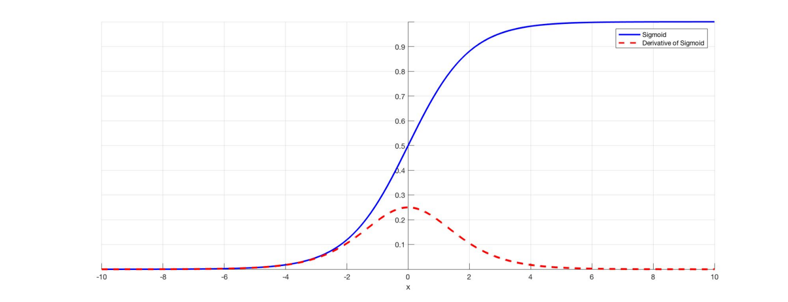
Optimizer: Adam (Adaptive Moment Estimation)

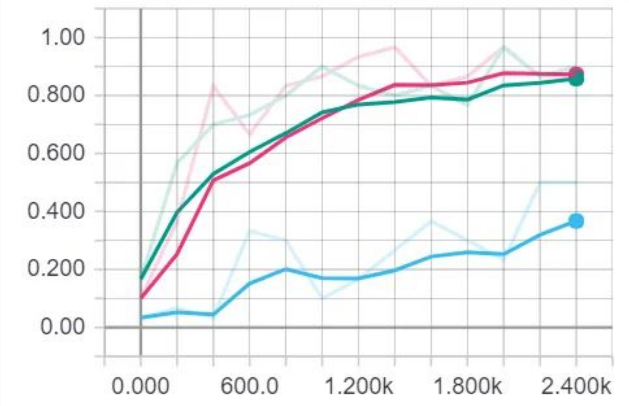
- Iteratively update the weights of the training model, in place of stochastic gradient descent.
- Individual learning rates for different parameters
- Requires minimal fine tuning of parameters



Activation: Sigmoid

- **Advantages**:
  - Range between 0 and 1: suitable for probability
  - Great for binary classification
- **Disadvantages**:
  - Vanishing gradient problem: slows down learning time for early layer parameters, inaccurate parameters

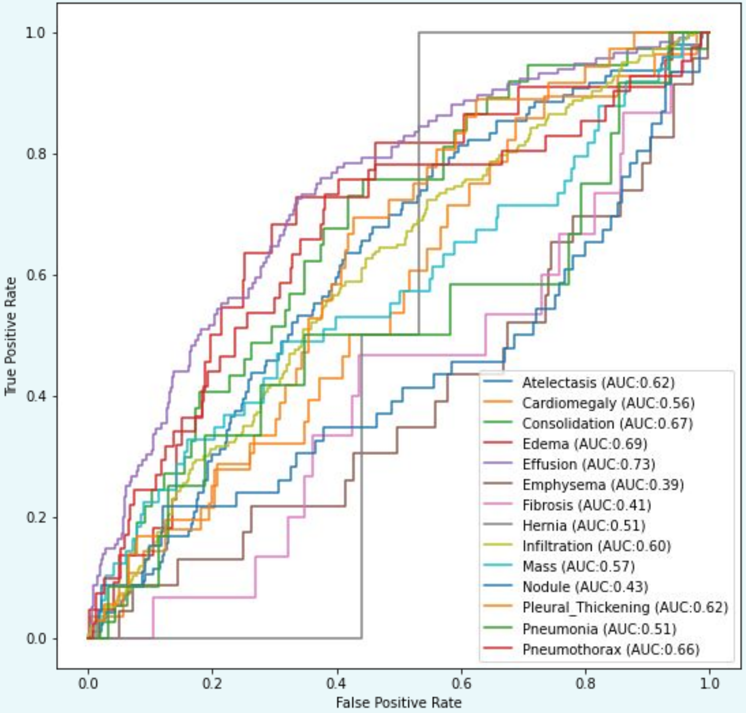




Accuracy of the three activation scenarios – sigmoid (blue), ReLU (red), Leaky ReLU (green)

RESULTS

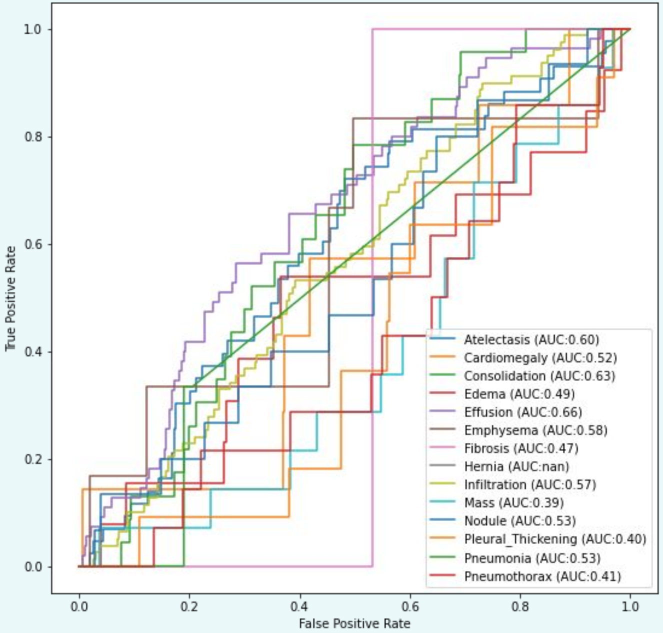
GENERAL



Contains 1402 validation images

```
Atelectasis: 9.38%
Cardiomegaly: 2.73%
Consolidation: 3.61%
Edema: 2.15%
Effusion: 11.33%
Emphysema: 2.25%
Fibrosis: 1.46%
Hernia: 0.20%
Infiltration: 17.29%
Mass: 4.79%
Nodule: 4.49%
Pleural_Thickening: 3.52%
Pneumonia: 1.17%
Pneumothorax: 4.00%
```

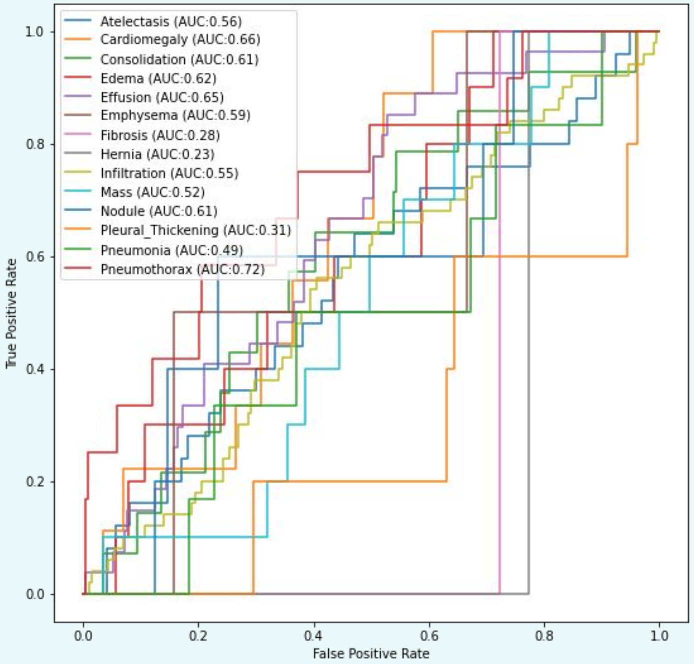
MALE & ANTERIOR-POSTERIOR



Contains 339 validation images

```
Atelectasis: 12.68%
Cardiomegaly: 2.06%
Consolidation: 6.78%
Edema: 3.83%
Effusion: 16.22%
Emphysema: 1.77%
Fibrosis: 0.29%
Hernia: 0.00%
Infiltration: 23.30%
Mass: 4.13%
Nodule: 4.42%
Pleural_Thickening: 3.24%
Pneumonia: 0.88%
Pneumothorax: 4.13%
```

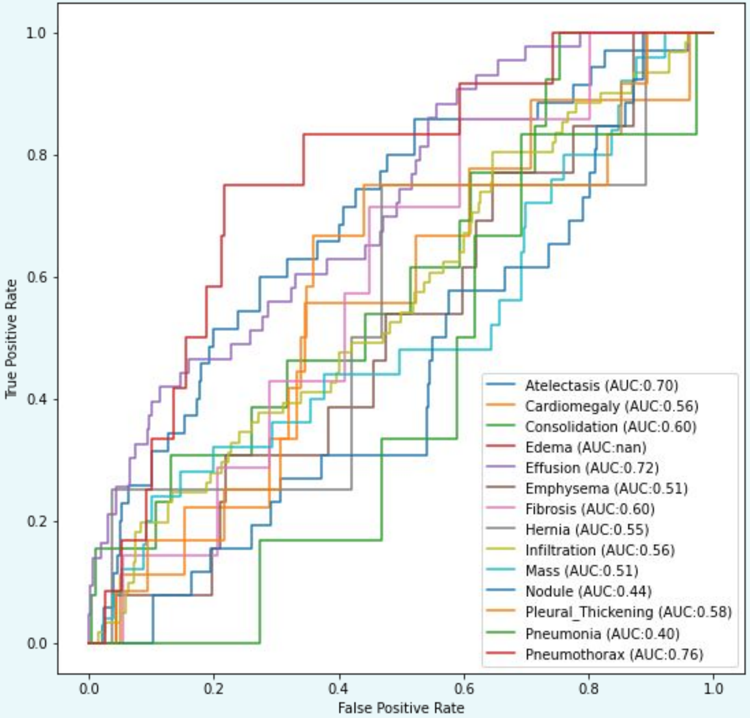
FEMALE & ANTERIOR-POSTERIOR



Contains 235 validation images

```
Atelectasis: 10.64%
Cardiomegaly: 3.83%
Consolidation: 5.96%
Edema: 4.26%
Effusion: 11.49%
Emphysema: 0.85%
Fibrosis: 0.43%
Hernia: 0.43%
Infiltration: 21.28%
Mass: 4.26%
Nodule: 2.13%
Pleural_Thickening: 2.13%
Pneumonia: 2.55%
Pneumothorax: 5.11%
```

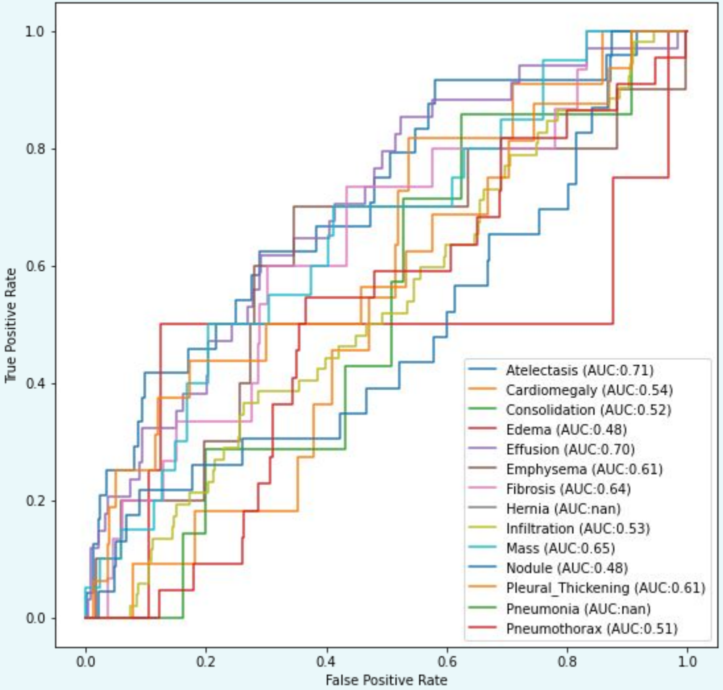
MALE & POSTERIOR-ANTERIOR



Contains 437 validation images

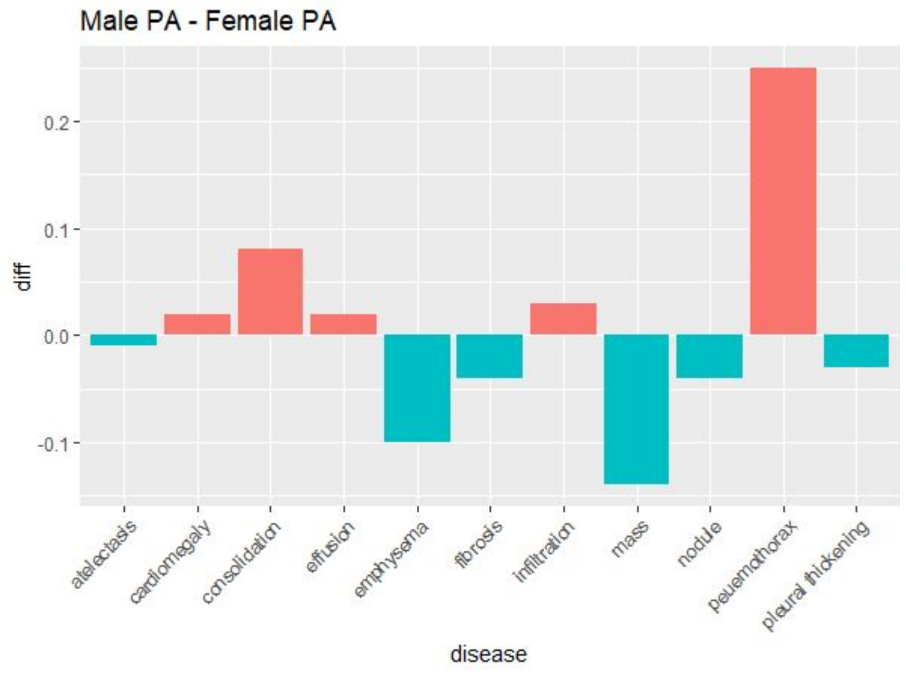
```
Atelectasis: 8.01%
Cardiomegaly: 2.06%
Consolidation: 2.97%
Edema: 0.00%
Effusion: 9.84%
Emphysema: 2.97%
Fibrosis: 1.60%
Hernia: 0.92%
Infiltration: 13.96%
Mass: 5.72%
Nodule: 5.95%
Pleural_Thickening: 2.75%
Pneumonia: 1.37%
Pneumothorax: 2.75%
```

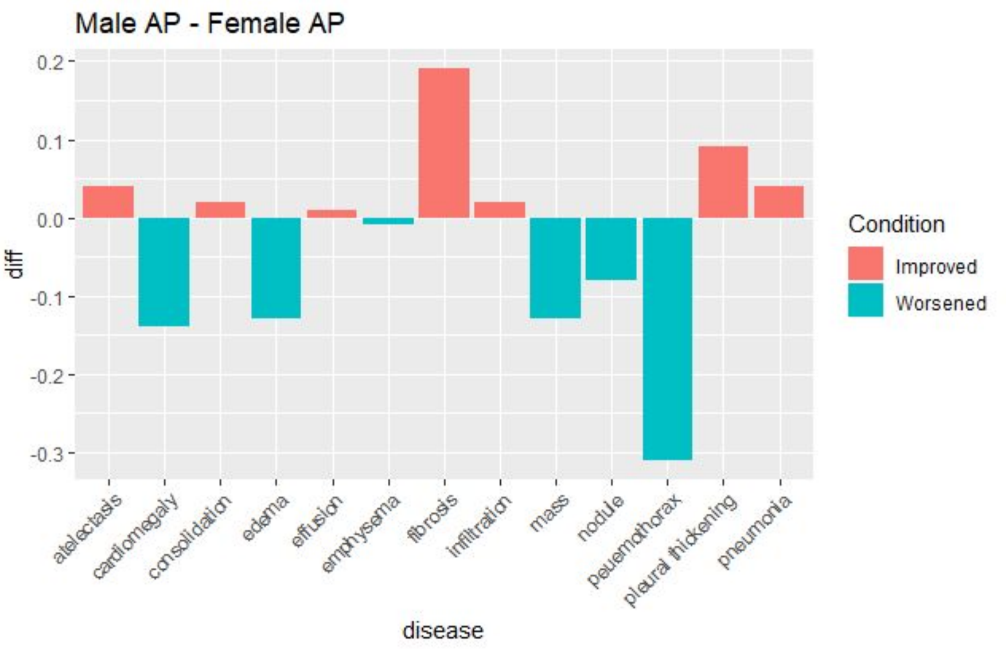
FEMALE & POSTERIOR-ANTERIOR

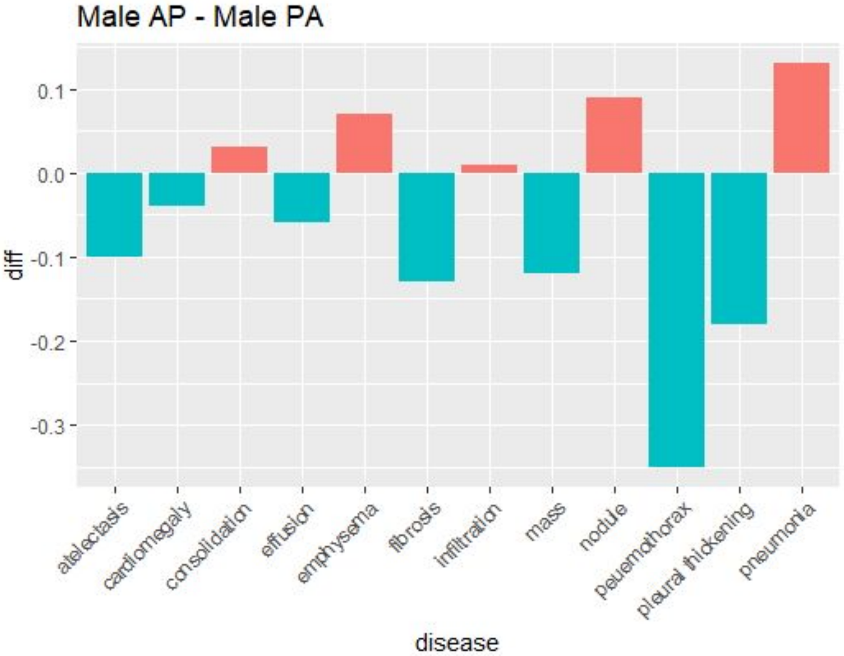


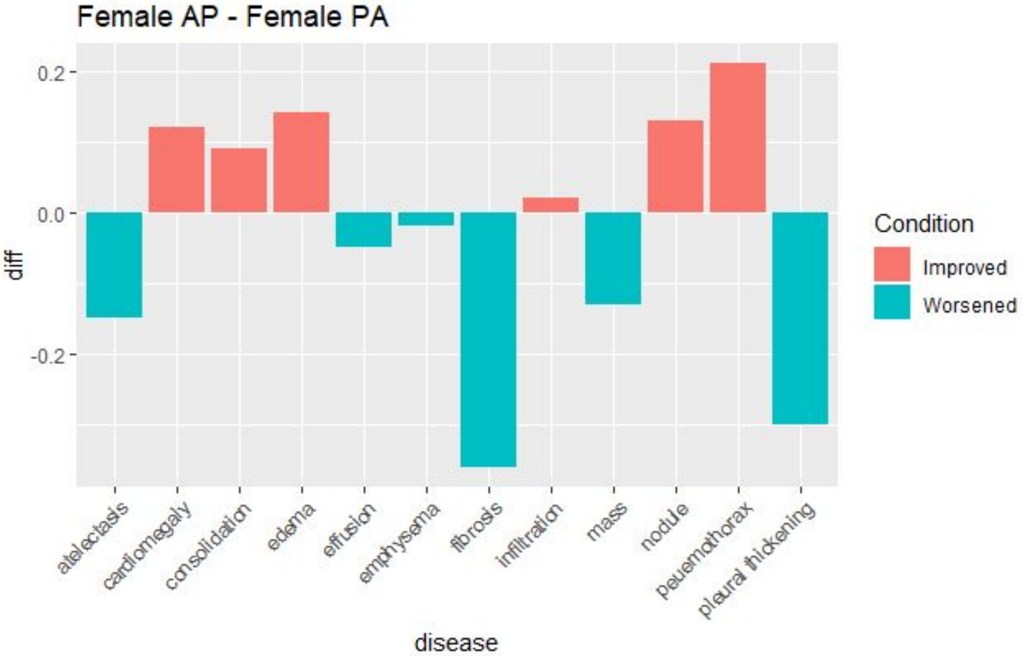
Contains 391 validation images

```
Atelectasis: 6.14%
Cardiomegaly: 2.81%
Consolidation: 1.79%
Edema: 1.02%
Effusion: 8.70%
Emphysema: 2.56%
Fibrosis: 3.84%
Hernia: 0.00%
Infiltration: 13.30%
Mass: 5.12%
Nodule: 5.88%
Pleural_Thickening: 4.09%
Pneumonia: 0.00%
Pneumothorax: 5.63%
```


Male PA - Female PA
0.2
0.1
0.0
-0.1
diff
atelectasis
cardiomegaly
consolidation
effusion
emphysema
fibrosis
infiltration
mass
nodule
peuemothorax
pleural thickening
disease


Male AP - Female AP
0.2
0.1
0.0
-0.1
-0.2
-0.3
diff
atelectasis
cardiomegaly
consolidation
edema
effusion
emphysema
fibrosis
infiltration
mass
nodule
peuemothorax
pleural thickening
pneumonia
disease
Condition
Improved
Worsened


Male AP - Male PA
0.1
0.0
-0.1
-0.2
-0.3
diff
atelectasis
cardiomegaly
consolidation
effusion
emphysema
fibrosis
infiltration
mass
nodule
peuemothorax
pleural thickening
pneumonia
disease


Female AP - Female PA
0.2
0.0
-0.2
diff
atelectasis
cardiomegaly
consolidation
edema
effusion
emphysema
fibrosis
infiltration
mass
nodule
peuemothorax
pleural thickening
disease
Condition
Improved
Worsened

CONCLUSIONS & FUTURE IMPROVEMENTS

From the results we have based on the sample dataset, our predictions based on sex and image view orientation did not show up well in the results. This might be due to the general inaccuracy of the models, since the nnets were running on a few hundred images.

For our future step, we hope to run our models on the full dataset with GPU. Currently, we are encountering difficulties with downloading and modifying the full dataset, but we hope to solve the problems soon.

THANKS

Does anyone have any questions?