FDA Submission

Your Name: Dinesh K.

Name of your Device: Pneumonia Reader Assistant

Algorithm Description

1. General Information

Intended Use Statement: Assist radiologist in classifying a given chest x-ray for the presence or absence of pneumonia.

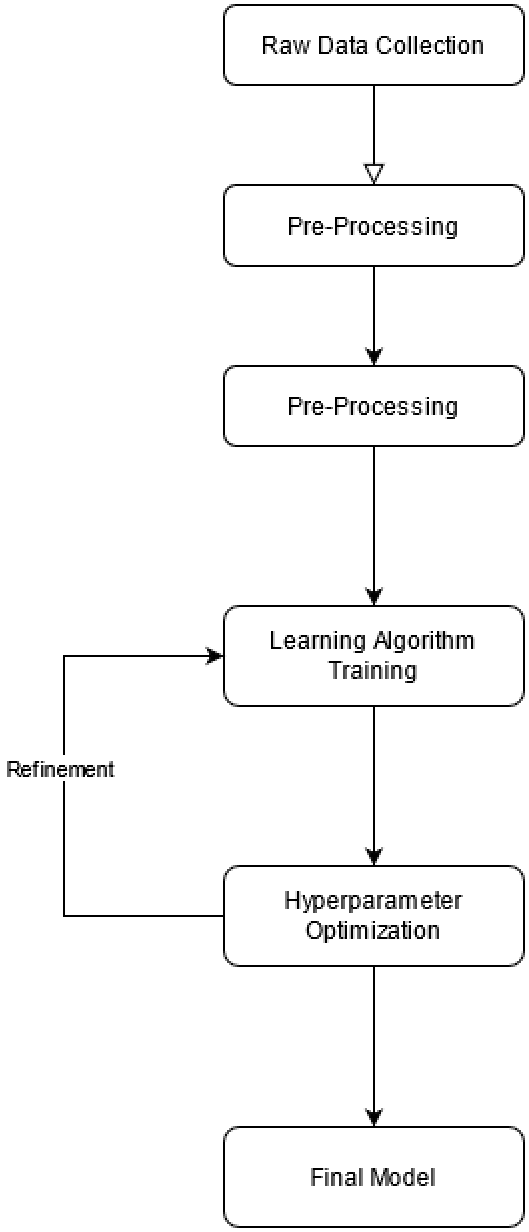
Indications for Use: Indicated for use in screening Pneumonia in males and females of ages 1-95 years having one or more combination of following diseases: atelectasis, heart enlargement, standardization, edema, effusion, emphysema, fibrosis, hernia, infiltration, mass, Creed, pleura thickening and pneumothorax.

Device Limitations:

- GPU processing required for running model. Otherwise, there will be time-lag in prediction.
- Model is trained on available NIH Chest x-ray dataset. The image intensity distribution of pneumonia and other diseases such as Infiltration and effusion are almost similar. This may lead to false results.

Clinical Impact of Performance: The algorithm is designed to have high recall. When a high recall test returns a negative result, we can be confident that the result is truly negative since a high recall test has low false negatives. This is especially useful in worklist prioritization where we want to make sure that people without the disease are being de-prioritized.

2. Algorithm Design and Function



DICOM Checking Steps: The DICOM metadata is checked for following:

- Modality = DX
- Body Part Examined = Chest
- Patient Position in `[PA, AP]`

Preprocessing Steps: All the images have been normalized in range of [0, 1] using standardization and then the images are resized to (224, 224, 3).

CNN Architecture: Model is created using a pre-trained VGG-16 network with fine tuning done on the last convolution layer. First 17 layers from VGG16 network was used as-is (Frozen /non-trainable). Model architecture is as follows,

```
Pneumonia_Classifier.summary()
```

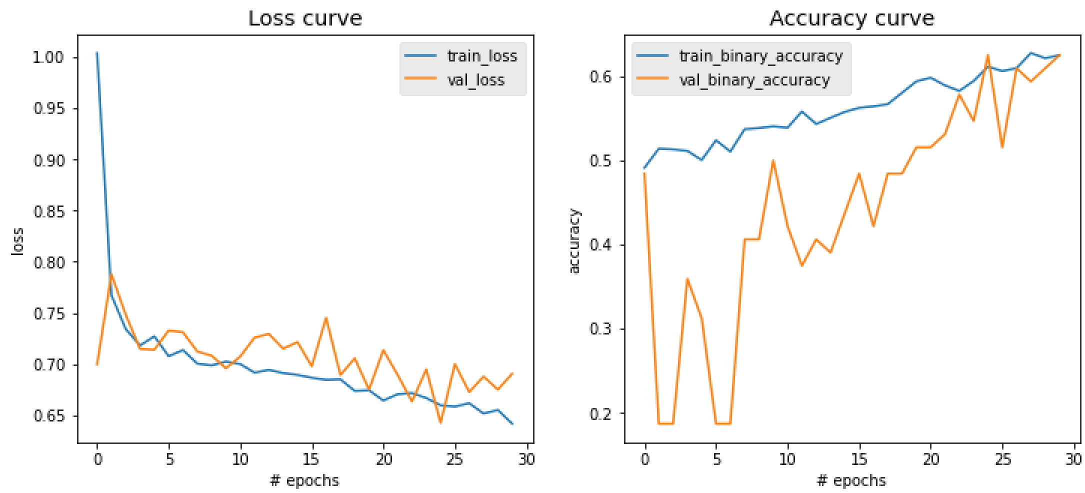
```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
model_1 (Model)              (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              25691136
_________________________________________________________________
dropout_2 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 512)               524800
_________________________________________________________________
dropout_3 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_3 (Dense)              (None, 256)               131328
_________________________________________________________________
dropout_4 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_4 (Dense)              (None, 64)                16448
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 65
=================================================================
Total params: 41,078,465
Trainable params: 28,723,585
Non-trainable params: 12,354,880
_________________________________________________________________
```

3. Algorithm Training

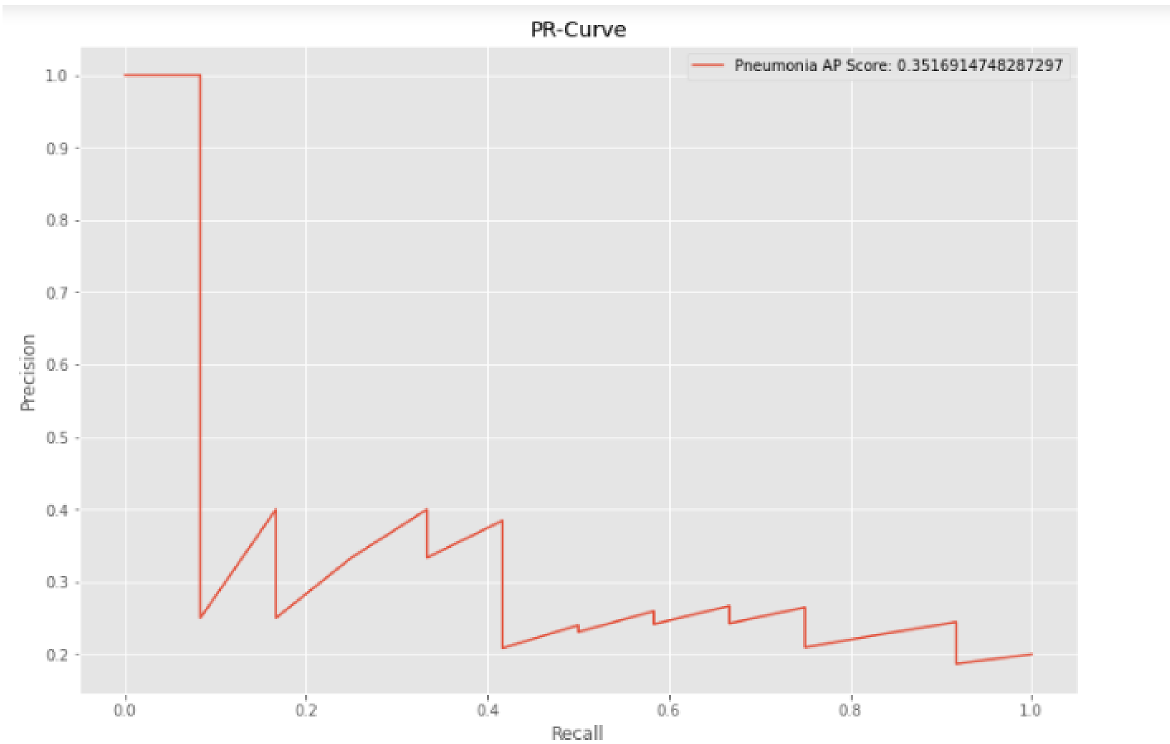
Parameters:

- Types of augmentation used during training:
    - Horizontal Flip
    - Random Rotation up to 20 degrees
    - Random shear range of (+/-) 10%
    - Random width shift of (+/-) 10%
    - Random height shift of (+/-) 10%
    - Random zoom range of (+/-) 12%
- Batch size: 128
- Optimizer learning rate: 1e-4

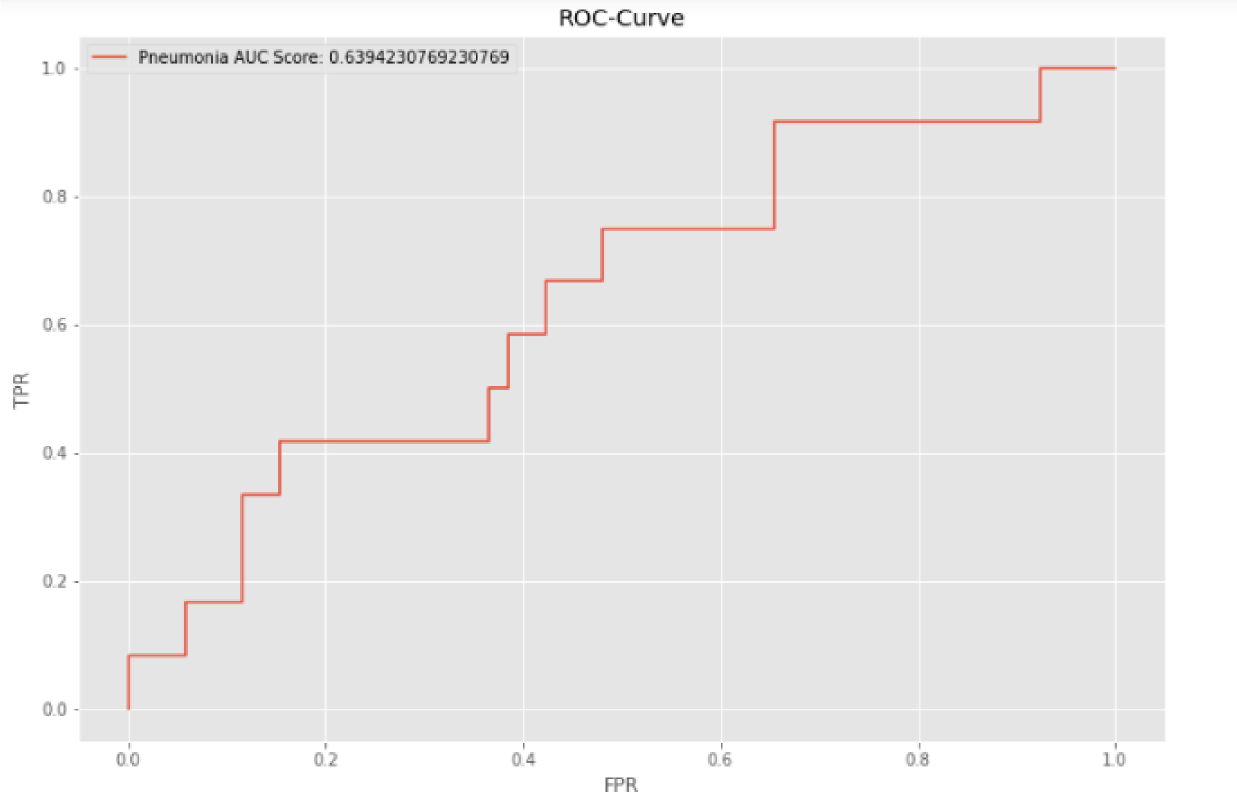
- Layers of pre-existing architecture that were frozen: First 17 layers of VGG pre-trained network
- Layers of pre-existing architecture those were fine-tuned: Last convolutional layer and all fully-connected layers of the architecture. Drop out of 0.5 is used during fully connected layer to fine tune model and avoid overfitting.
- **Loss and Accuracy curve**



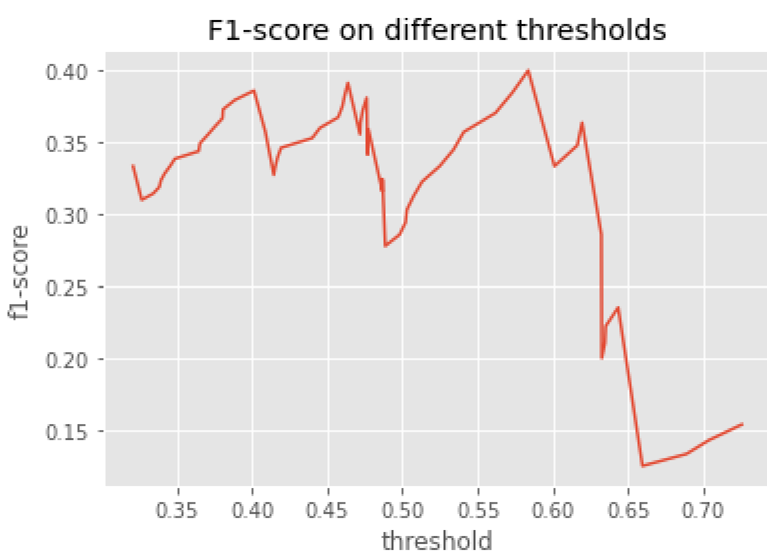
- **Precision-Recall Curve**



- **Precision-Recall Curve**



- **F1 Scores on different thresholds**



```
Precision at 0.5836243033409119 threshold is 0.38461538461538464.
Recall at 0.5836243033409119 threshold is 0.4166666666666667.
F1 score at 0.5836243033409119 threshold is 0.4.
```

Final Threshold and Explanation:

- Maximum F1 score of this model at threshold of 0.58 is **0.4**.
- This is slightly higher than average Radiologist F1 score which is at 0.38 as per ChexNet paper (https://arxiv.org/pdf/1711.05225.pdf).

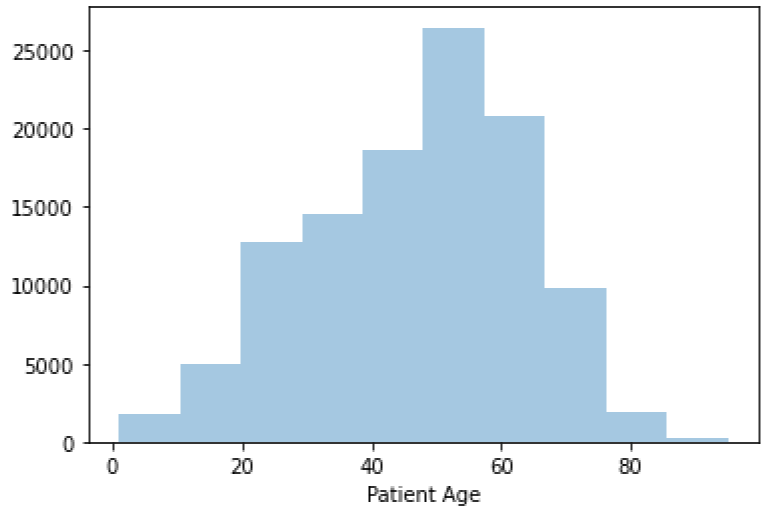
4. Databases

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging. The lack of large publicly available datasets with annotations means it is still very difficult, if not impossible, to achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites with chest X-rays.

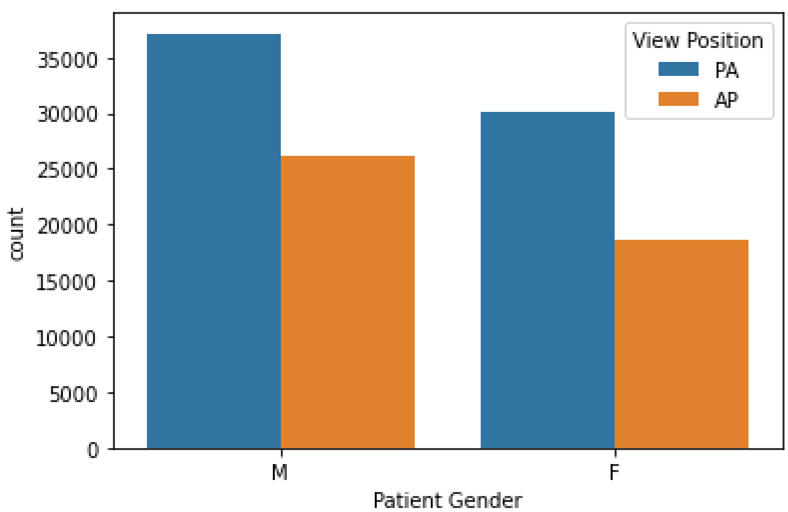
This NIH Chest X-ray Dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. The CSV file has following columns:

- Image Index: File name
- Finding Labels: Disease type (Class label)
- Follow-up #
- Patient ID
- Patient Age
- Patient Gender
- View Position: X-ray orientation
- OriginalImageWidth
- OriginalImageHeight
- OriginalImagePixelSpacing_x
- OriginalImagePixelSpacing_y

- Patient Age Distribution

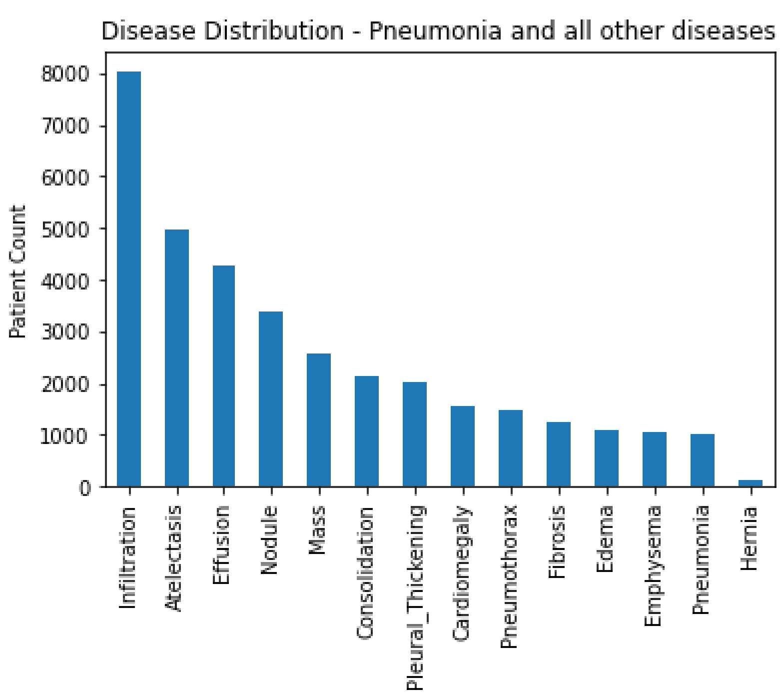


- Patient Gender Distribution along with X-ray view position (PA vs AP)

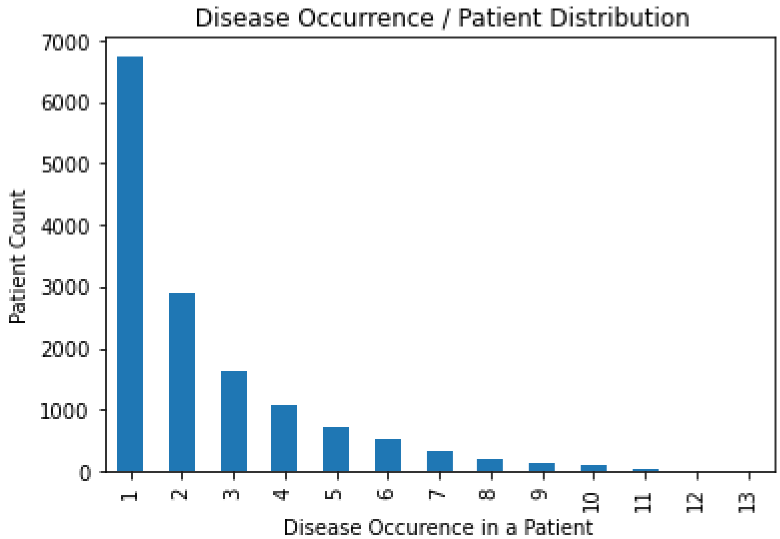


- Pneumonia patients – Age, Gender and X-ray view position (PA vs AP) Distribution

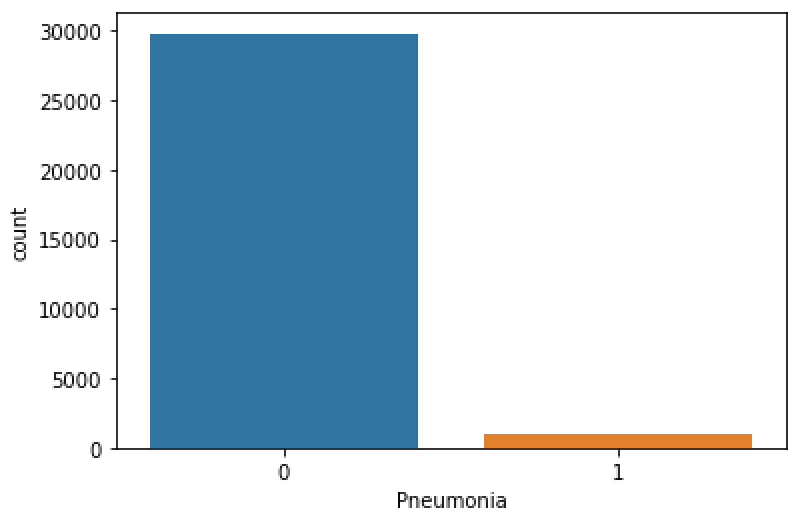
- Disease Distribution - Pneumonia and all other diseases



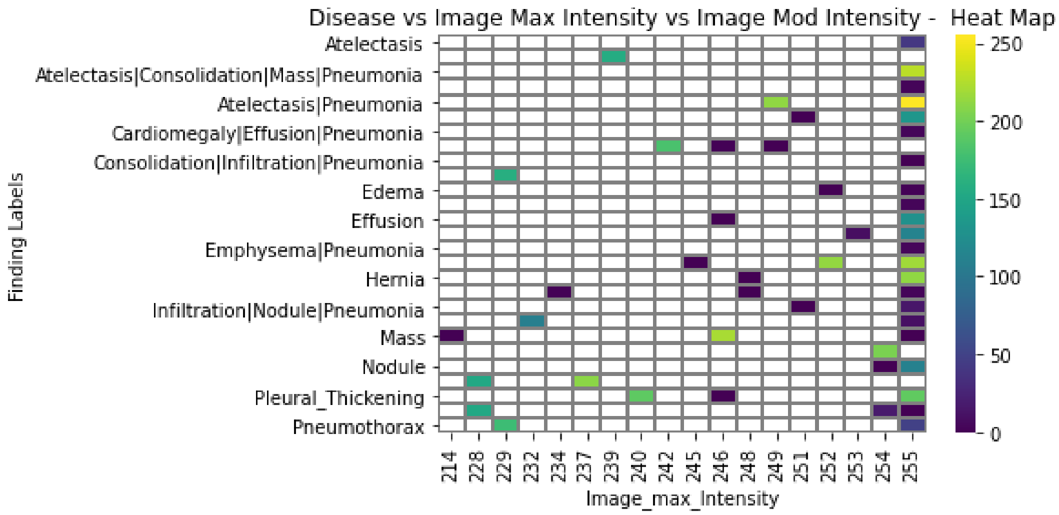
- Disease Occurrence/Patient Distribution - Pneumonia and all other diseases

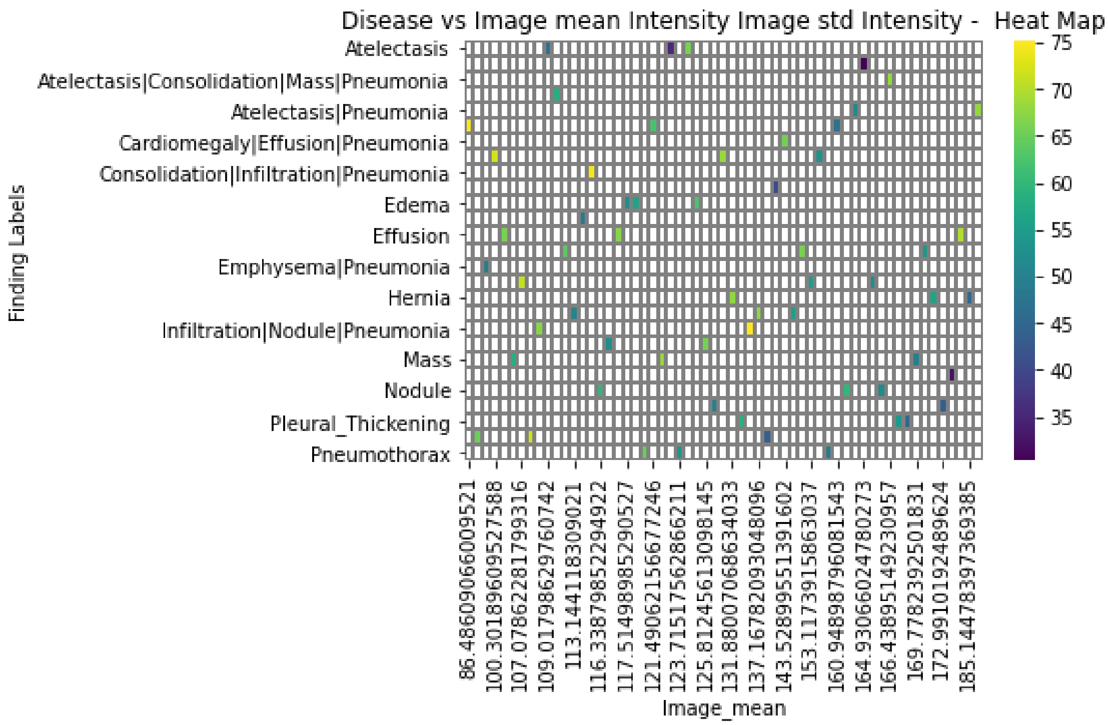


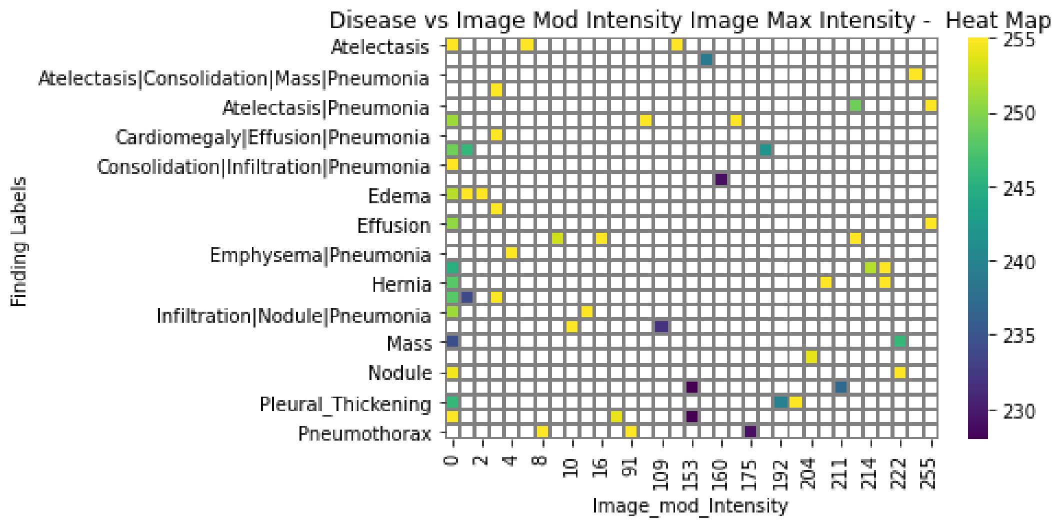
- Patient count with Pneumonia(1) and Non-Pneumonia disease(0)

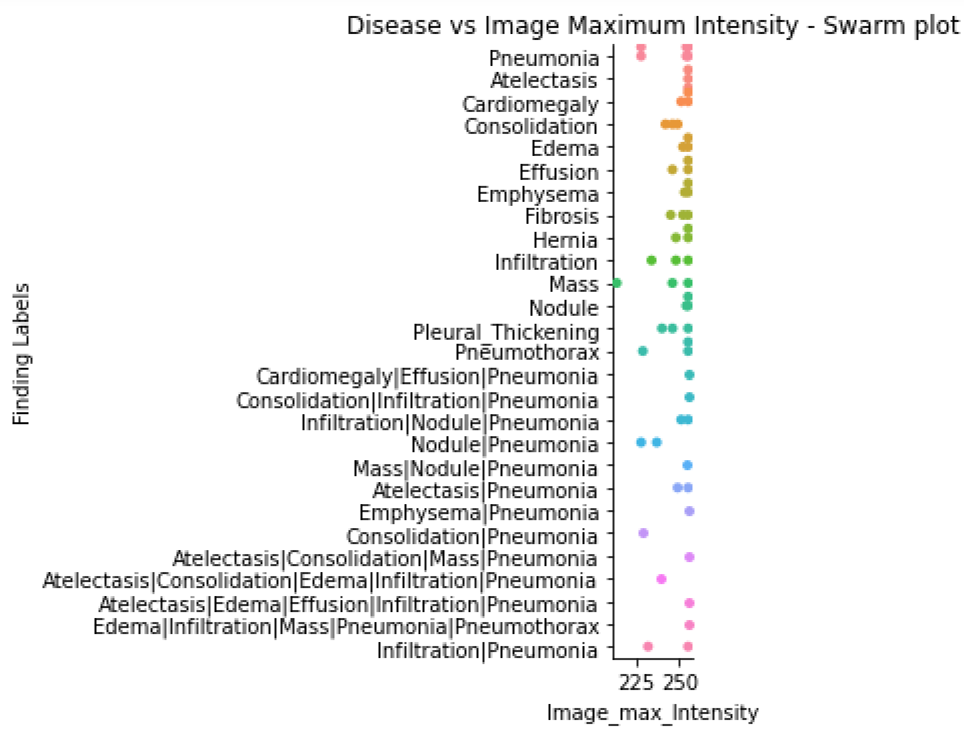


- Diseases pixel intensity plots(Sample data)

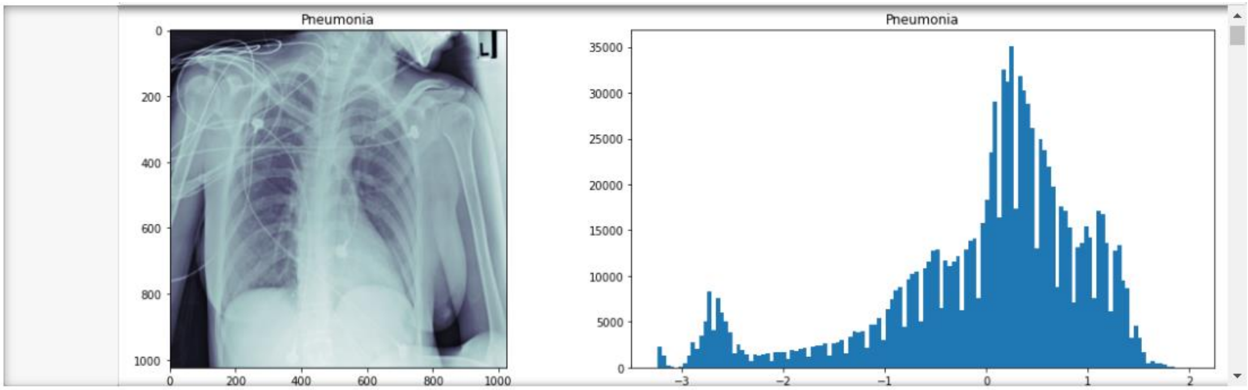

Disease vs Image Max Intensity vs Image Mod Intensity - Heat Map
Atelectasis
Atelectasis|Consolidation|Mass|Pneumonia
Atelectasis|Pneumonia
Cardiomegaly|Effusion|Pneumonia
Consolidation|Infiltration|Pneumonia
Edema
Effusion
Emphysema|Pneumonia
Hernia
Infiltration|Nodule|Pneumonia
Mass
Nodule
Pleural_Thickening
Pneumothorax
Finding Labels
214 228 229 232 234 237 239 240 242 245 246 248 249 251 252 253 254 255
Image_max_Intensity
0 50 100 150 200 250


Disease vs Image mean Intensity Image std Intensity - Heat Map
Atelectasis
Atelectasis|Consolidation|Mass|Pneumonia
Atelectasis|Pneumonia
Cardiomegaly|Effusion|Pneumonia
Consolidation|Infiltration|Pneumonia
Edema
Effusion
Emphysema|Pneumonia
Hernia
Infiltration|Nodule|Pneumonia
Mass
Nodule
Pleural_Thickening
Pneumothorax
Finding Labels
86.48609066009521
100.3018960952758
107.07862281799316
109.01798629760742
113.144118309021
116.33879852294922
117.51498985290527
121.49062156677246
123.71517562866211
125.81245613098145
131.88007068634033
137.16782093048096
143.52899551391602
153.11739158630 37
160.94898796081543
164.93066024780273
166.43895149230957
169.77823925018 31
172.99101924896 24
185.14478397369385
Image_mean
35 40 45 50 55 60 65 70 75





- Sample Pneumonia image and standardized intensity distribution



Description of Training Dataset: The training data has 2290 chest-xray images. These are evenly distributed to have 50% pneumonia cases and 50% non-pneumonia cases.

```
Total number of records in this dataset is: 112120
###Training Dataset Information### :
Total number of records in the training dataset is: 2290
Total number of Pneumonia records in the training dataset is: 1145
Total number of Non-Pneumonia records in the training dataset is: 1145
% of Pneumonia cases in the training dataset is: 50.0
% of Non-Pneumonia cases in the training dataset is: 50.0
```

Description of Validation Dataset: The training data has 1430 chest-x-ray images. These are distributed to have 20% pneumonia cases and 80% non-pneumonia cases to be a representation of real-world scenario.

```
Total number of records in this dataset is: 112120
### Validation Dataset Information###:
Total number of records in the validation dataset is: 1430
Total number of Pneumonia records in the validation dataset is: 286
Total number of Non-Pneumonia records in the validation dataset is: 1144
% of Pneumonia cases in the validation dataset is: 20.0
% of Non-Pneumonia cases in the validation dataset is: 80.0
```

5. Ground Truth

To create these ground truth, the authors used Natural Language Processing to text-mine disease classifications from the associated radiological reports. The labels are expected to be >90% accurate and suitable for weakly-supervised learning.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: The FDA Validation dataset contains patients (both male and female) in age-range of 1-95 years. It includes 20% pneumonia cases and 80% non-pneumonia cases. Dataset includes multiple follow-up records for a given patient. Total number of records in the validation set is 1430.

Ground Truth Acquisition Methodology: The image labels are NLP extracted with estimated accuracy > 90 %. There could be chance that some labels are incorrect.

Algorithm Performance Standard: The algorithm performance has been validated using F1-score. The F1 score combines both precision and recall. F1 score allows us to better measure a test's accuracy when there are class imbalances. Model F1-score is 0.4 which is better than radiologist average F1-score of 0.387.