

FDA Submission

Your Name: Soya Aoki

Name of your Device: Pneumonia AutoScreener

Algorithm Description

1. General Information

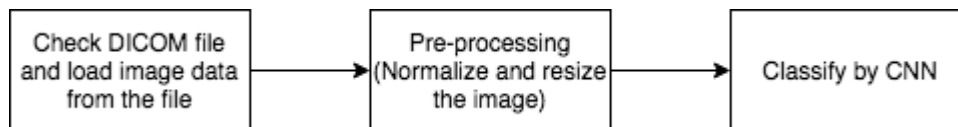
Intended Use Statement: Assisting a radiologist in classifying for the presence or absence of Pneumonia by chest x-ray image.

Indications for Use: Indicated for use in screening Pneumonia in males and females of ages 1-95 years.

Device Limitations: Requires Nvidia GPU. Otherwise user would face latency issues.

Clinical Impact of Performance: The algorithm is designed to have high recall. When a high recall algorithm returns a negative result, we can be confident that the result is truly negative since a high recall one has low false negatives. This is specially 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
- Patient Position in [PA, AP]

Pre-processing Steps: All the images have been normalized in range of [0, 1] and the images are resized to (224, 224, 3).

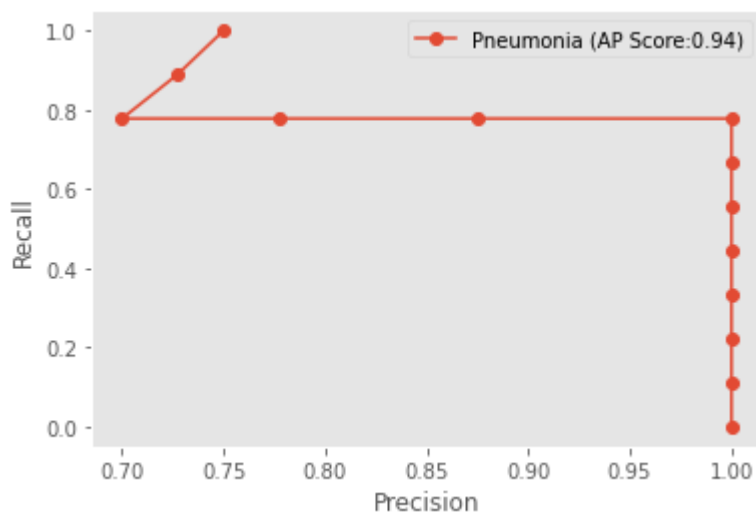
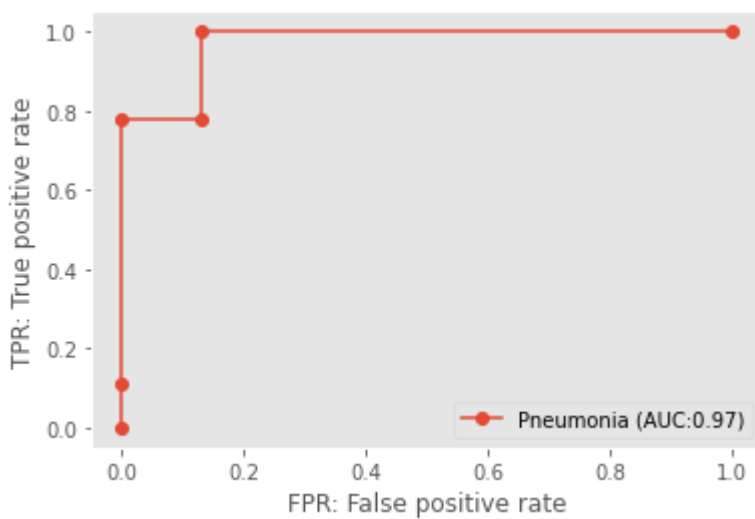
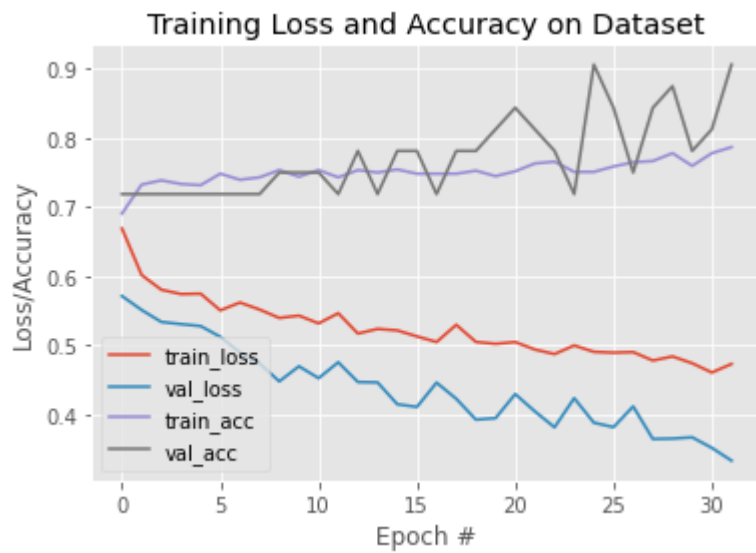
CNN Architecture: We have used a pre-trained VGG-16 network with fine tuning done on the last convolution layer. The architecture details are as follows:

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
dense_4 (Dense)	(None, 1)	257
Total params: 41,062,209		
Trainable params: 28,707,329		
Non-trainable params: 12,354,880		

3. Algorithm Training

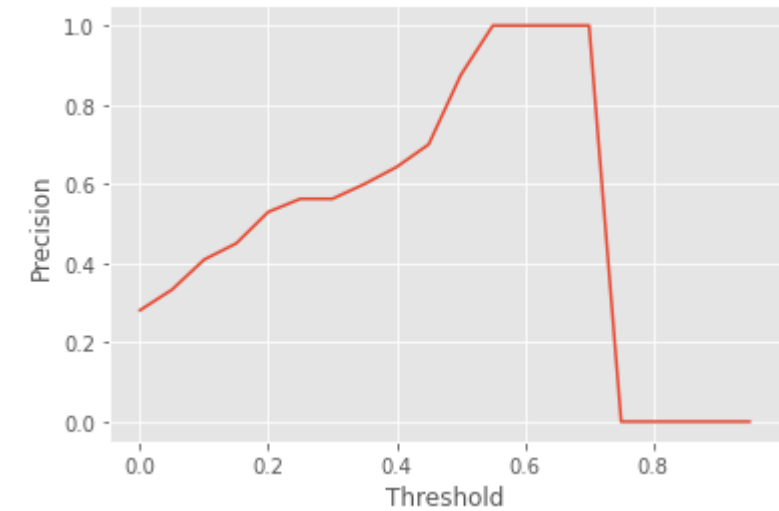
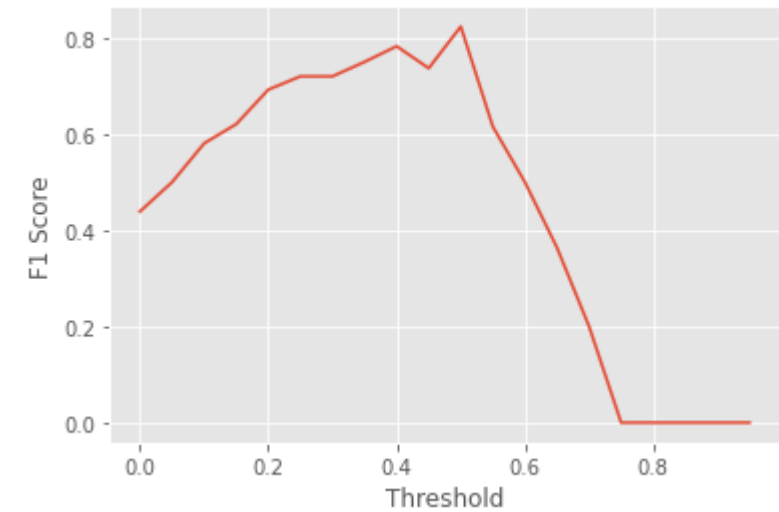
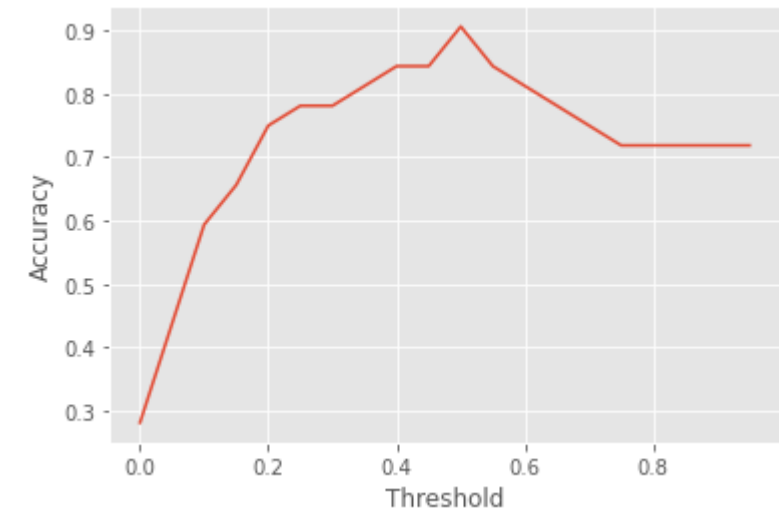
Parameters:

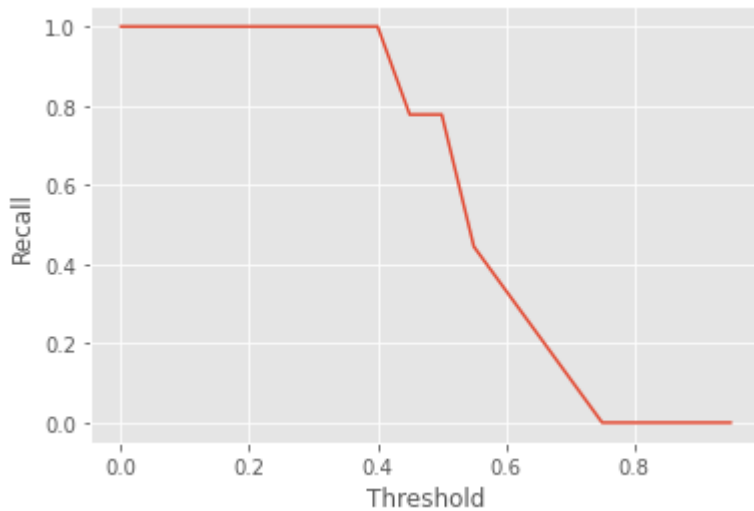
- Types of augmentation used during training:
 - Horizontal Flip
 - Random Rotation upto 20 degrees
 - Random shear range of (+/-) 10%
 - Random width shift of (+/-) 10%
 - Random height shift of (+/-) 10%
 - Random zoom range of (+/-) 10%
- Batch size: 32
- Optimizer and learning rate: Adam(lr=1e-4)
- Layers of pre-existing architecture that were frozen: First 17 layers of VGG pre-trained network
- Layers of pre-existing architecture that were fine-tuned: Last convolutional layer and all fully-connected layers of the architecture.



Final Threshold and Explanation:

The final threshold is chosen based on f1_score vs threshold curve. The threshold value that gives the maximum f1-score is chosen. It's 0.5 in our case.



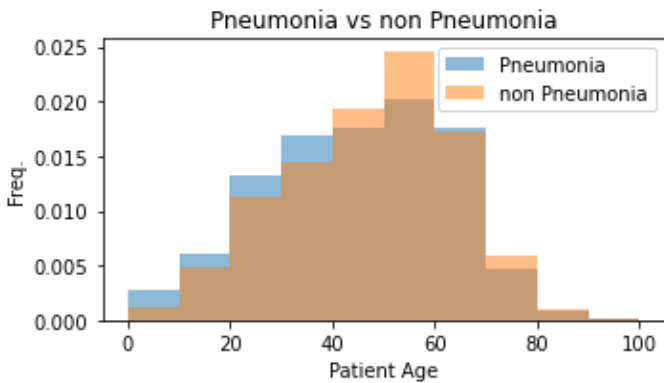
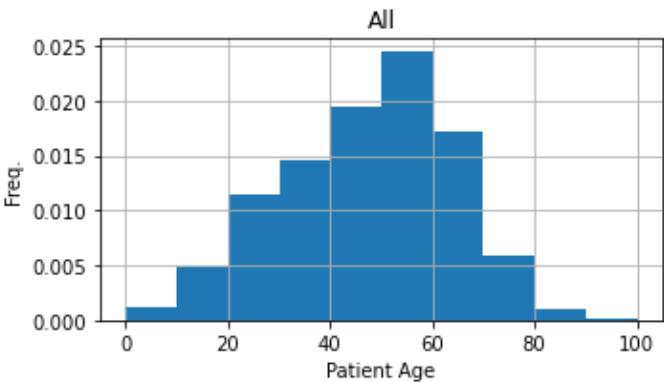
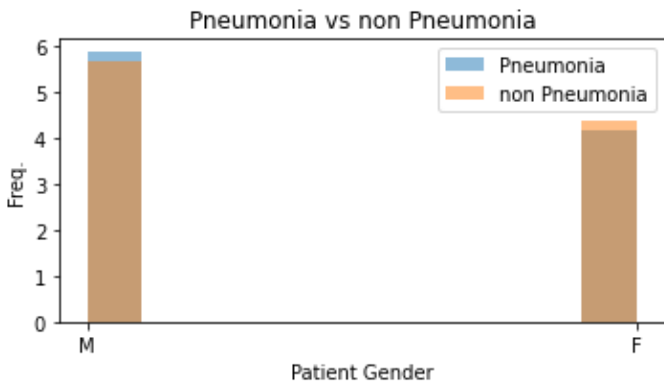
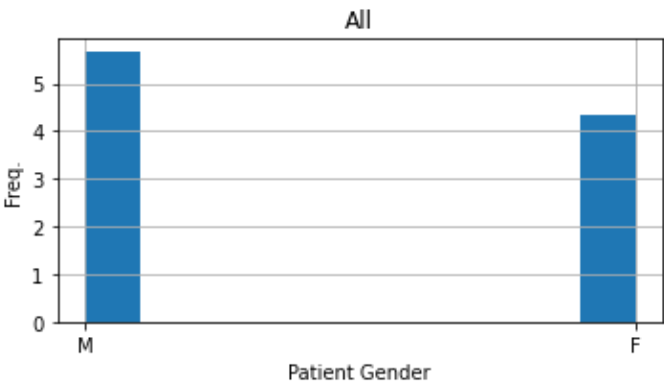
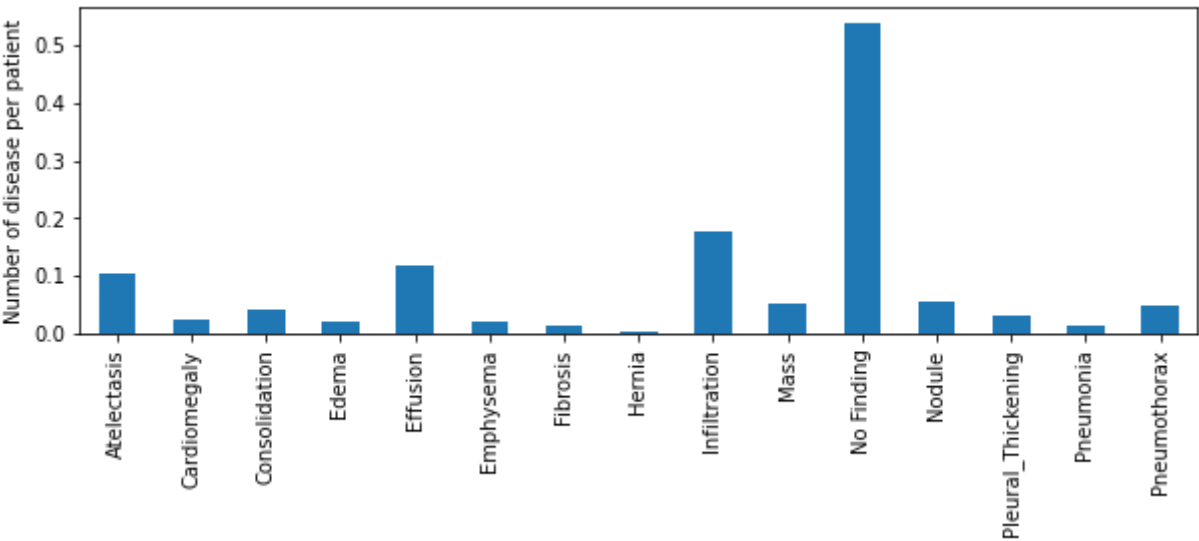
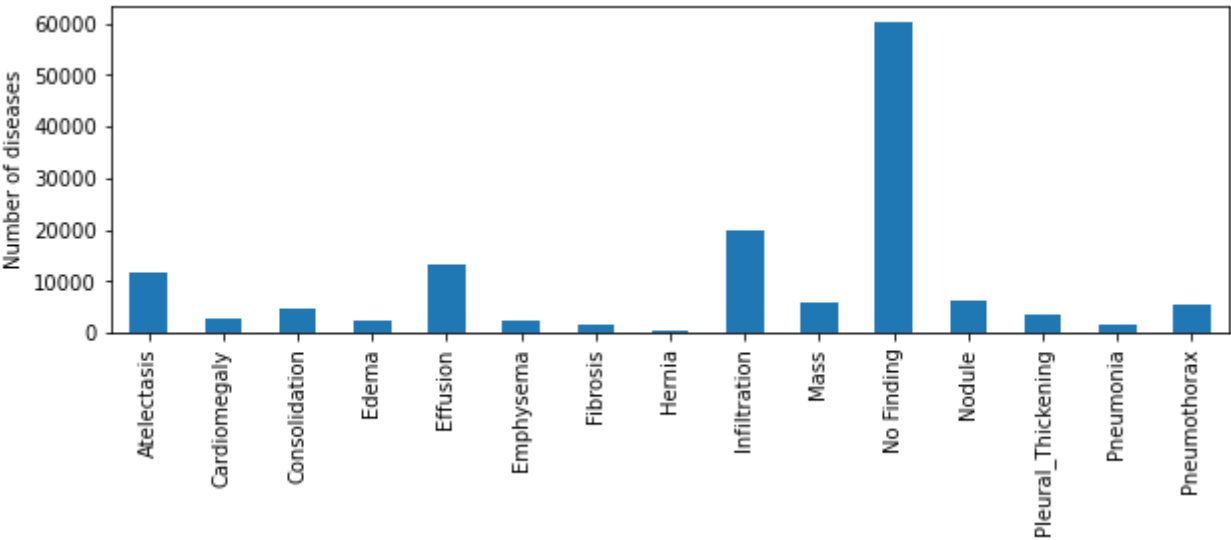


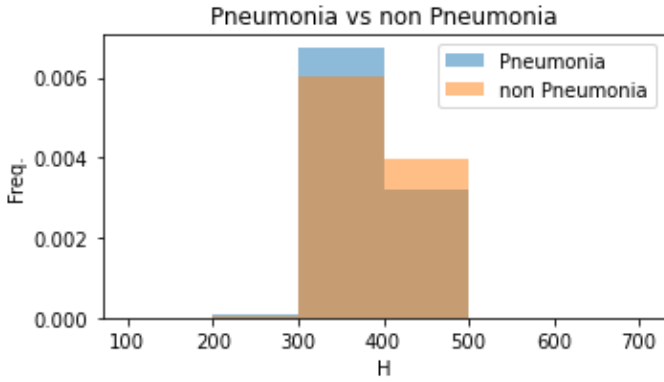
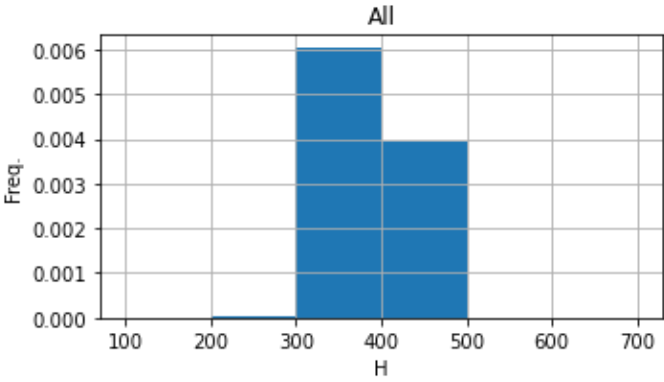
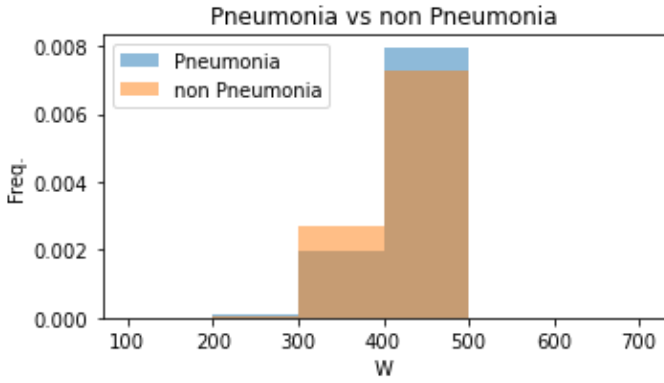
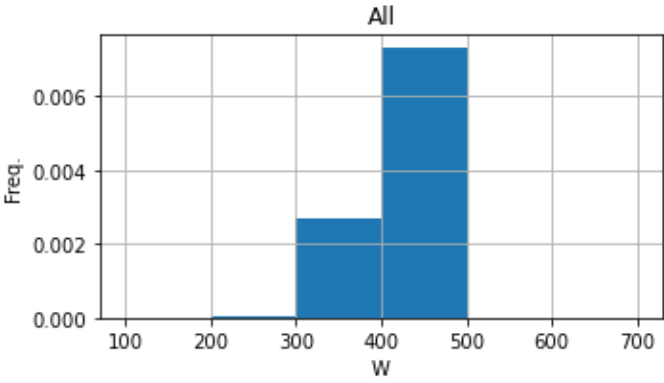
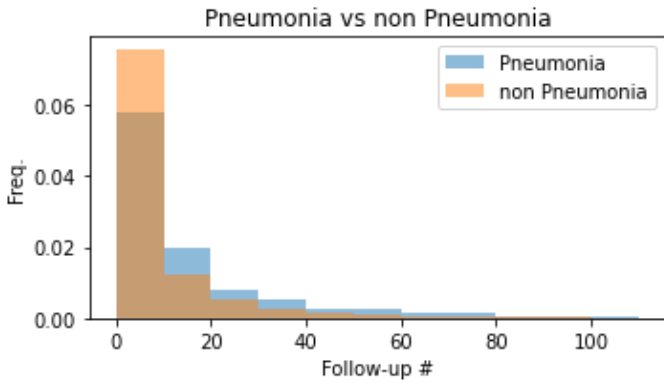
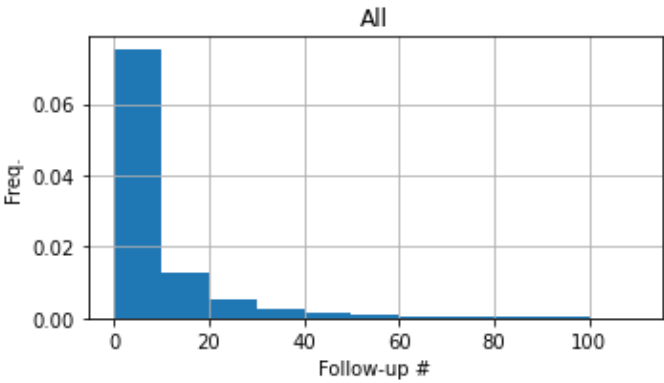
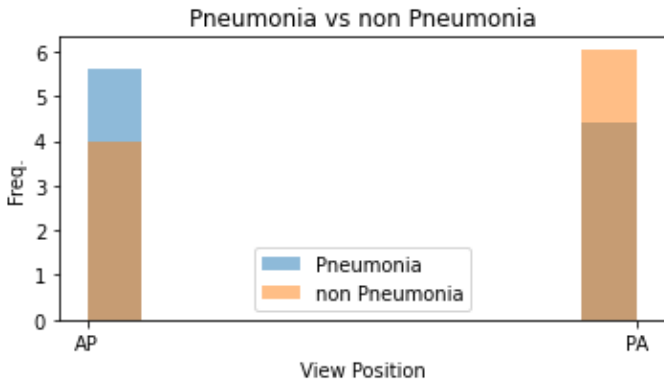
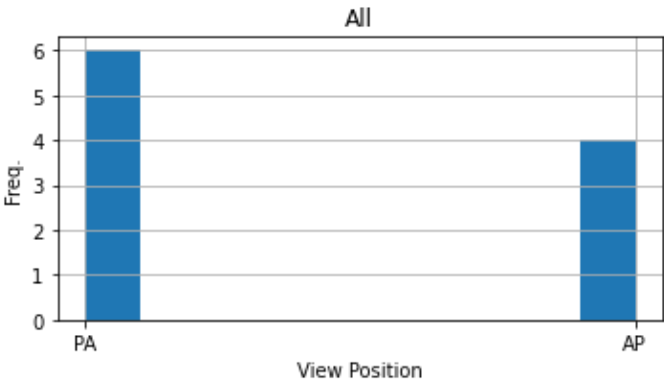
4. Databases

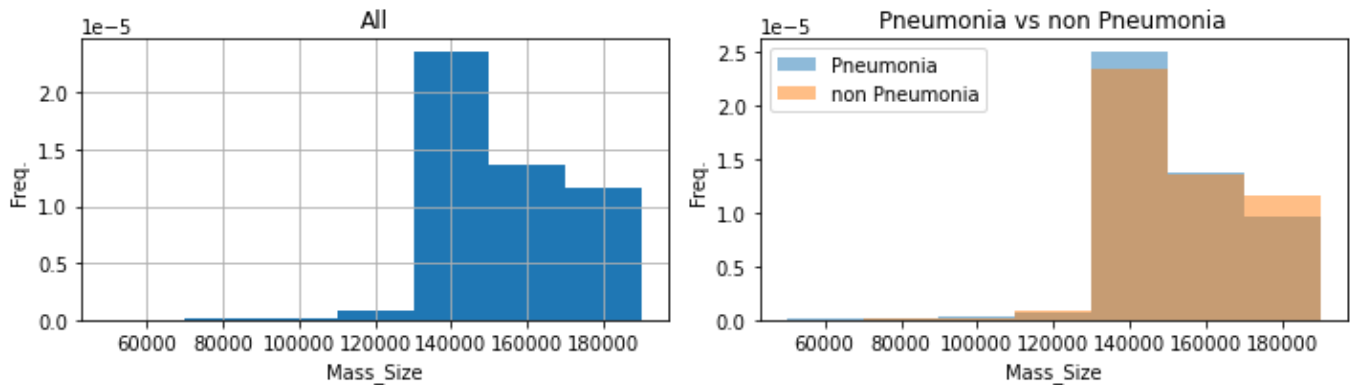
Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging. The lack of large publicly available datasets with annotations means 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







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.

Description of Validation Dataset: The training data has 1144 chest-xray images. These are distributed to have 25% pneumonia cases and 75% non-pneumonia cases.

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 25% pneumonia cases and 75% non-pneumonia cases. Further each patient have multiple follow-ups.

Ground Truth Acquisition Methodology: The image labels are NLP extracted so there could be some erroneous labels but the NLP labeling accuracy is estimated to be >90%.

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. We got an F1-score of 0.82 which is better than radiologist average F1-score of 0.387.