FDA Submission

Your Name: Tai Le

Name of your Device: Deep Learning Model To Detect Pneumonia From Chest X-Rays

Algorithm Description

1. General Information

Intended Use Statement:

The model is intended to classify a given chest x-ray for the presence or absence of pneumonia to assist radiologists in terms of prioritization and providing insights to possible findings by way of segmentation and localization visibility.

Indications for Use:

The model can be used on X-Rays , 2D medical images, with either AP or PA viewing position taken from human male or female patients between the ages of 1 and 100 years old.

Device Limitations:

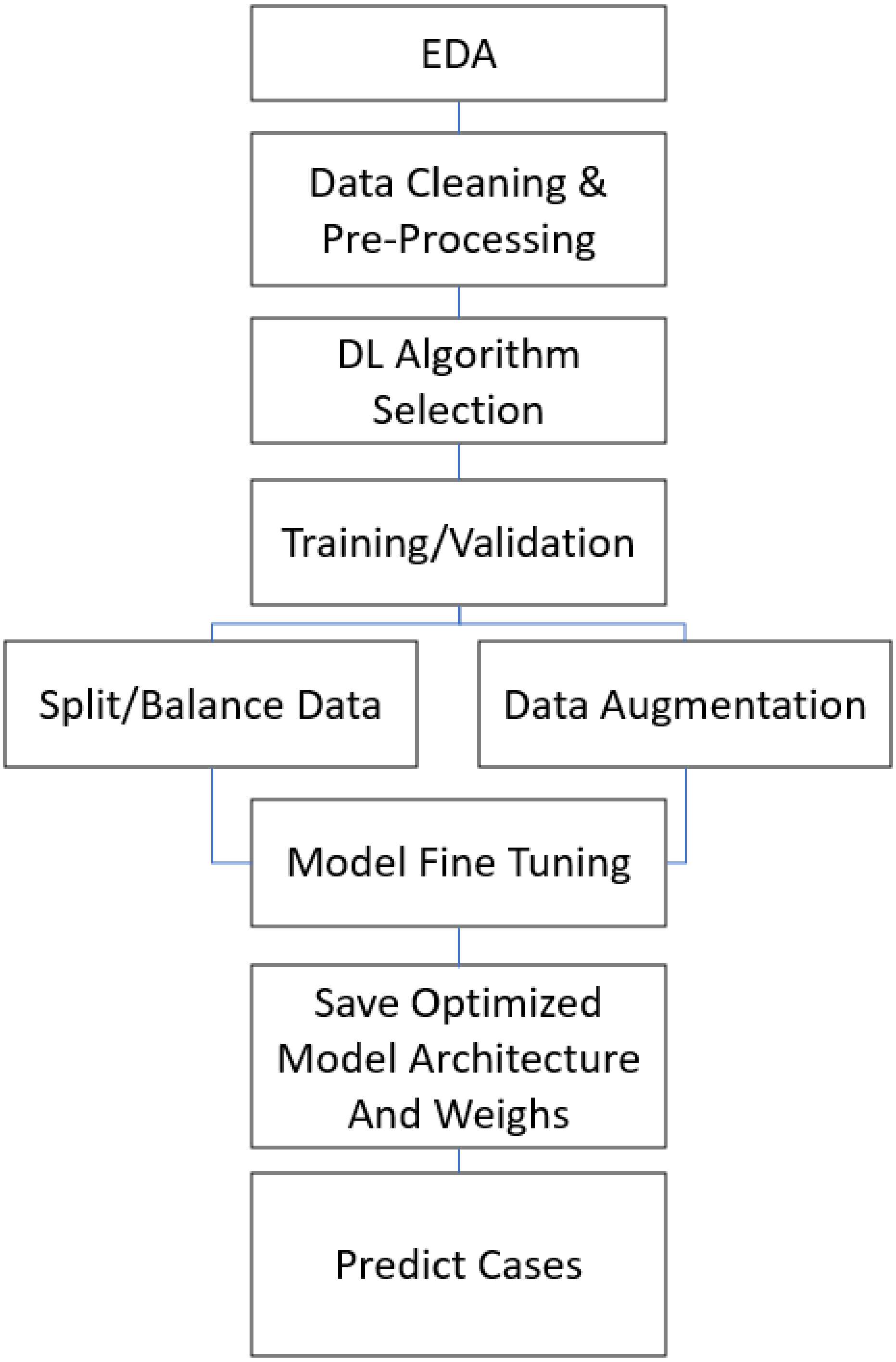
- The model performance is highly degraded without the sufficient use of computing infrastructure such as CPU/GPU, fast network I/O and RAM.
- The quality and sufficient amount of data used in training, validation and inference are important to the model classification.
- If other conditions such as infiltration and edema present, the model might incorrectly classify the X-ray with pneumonia presence.
- The model does not differentiate patient's with prior pneumonia history such emphasizing more weighs for past positive screenings.

Clinical Impact of Performance:

The intended use of the model to assist the radiologists in screening the X-rays. A false negative predicting the absence of pneumonia in a patient while he/she does have pneumonia is devasting. As with radiologists, the model is not perfect.

The appropriate use of the model, however, can improve the radiologist's workload. The final evaluation report of the patient remains in the control and responsibility of the radiologist.

2. Algorithm Design and Function



DICOM Checking Steps:

The model have these requirements for DICOM data:

- Chest X-Ray for body examined and with Modal "DX"
- Viewing position is "AP" or "PA"
- Patient age span between 1 and 100 years old

Preprocessing Steps:

- Split the dataset to training/validation using sklearn.model_selection library (80/20)
- Balance the split data, so the positive and negative pneumonia records are proportional
- Rescale and augment the images using Keras ImageDataGenerator to appropriately fit to VGG16 CNN
- Train and fine tune the model for optimization per specifications

CNN Architecture:

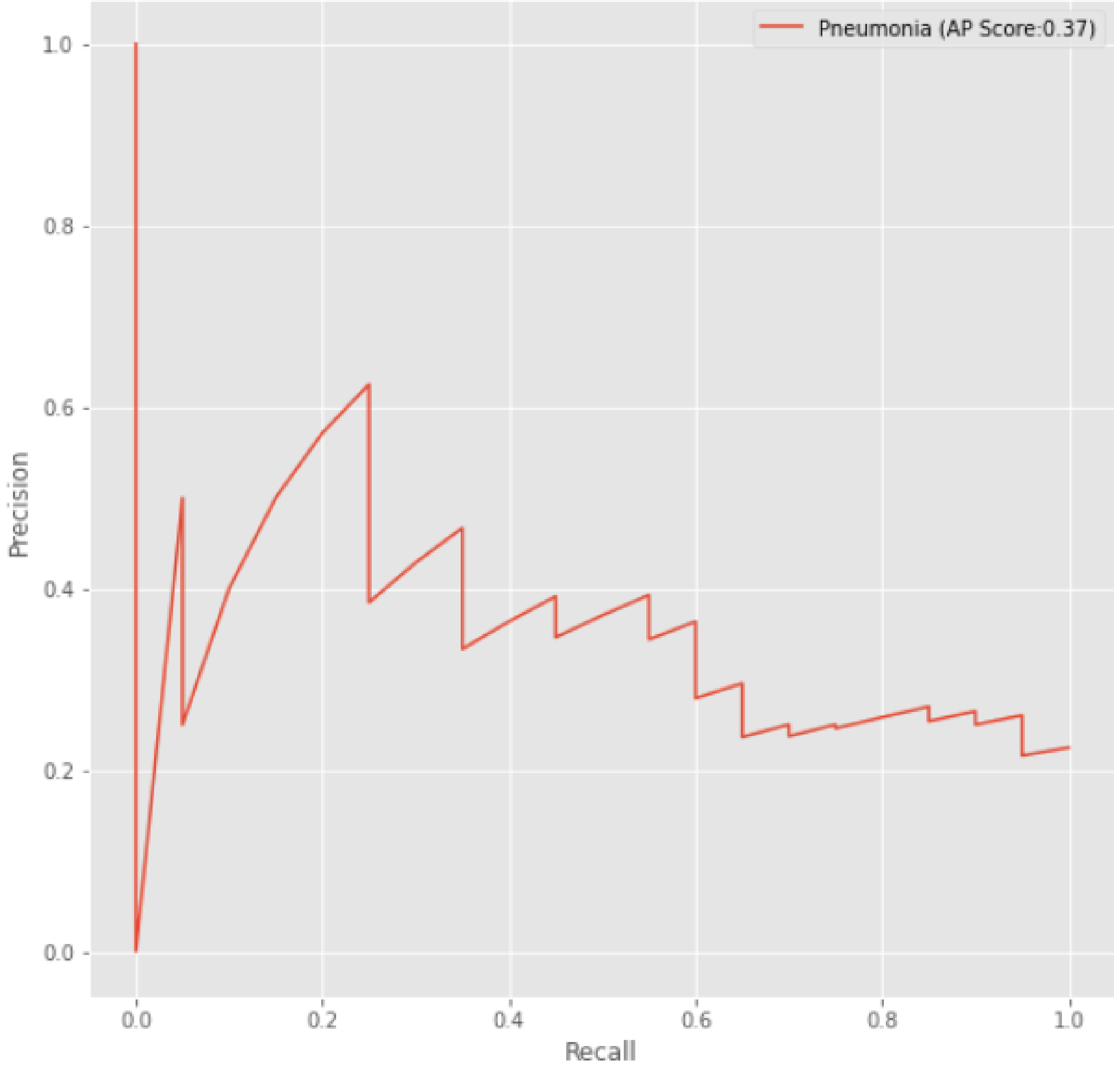
VGG16 CNN used for transfer learning with ImageNet weighs pretrained with over [1 million images].

3. Algorithm Training

Parameters:

- Types of augmentation used during training
    - rescale=1. / 255.0
    - horizontal_flip = True
    - vertical_flip = False
    - height_shift_range= 0.1
    - width_shift_range=0.1
    - rotation_range=20
    - shear_range = 0.1
    - zoom_range=0.1
- Batch size: 32
- Optimizer learning rate: Adam(learning_rate = 1e-4)
- Layers of pre-existing architecture that were frozen: First 17 layers
- Layers of pre-existing architecture that were fine-tuned: None
- Layers added to pre-existing architecture

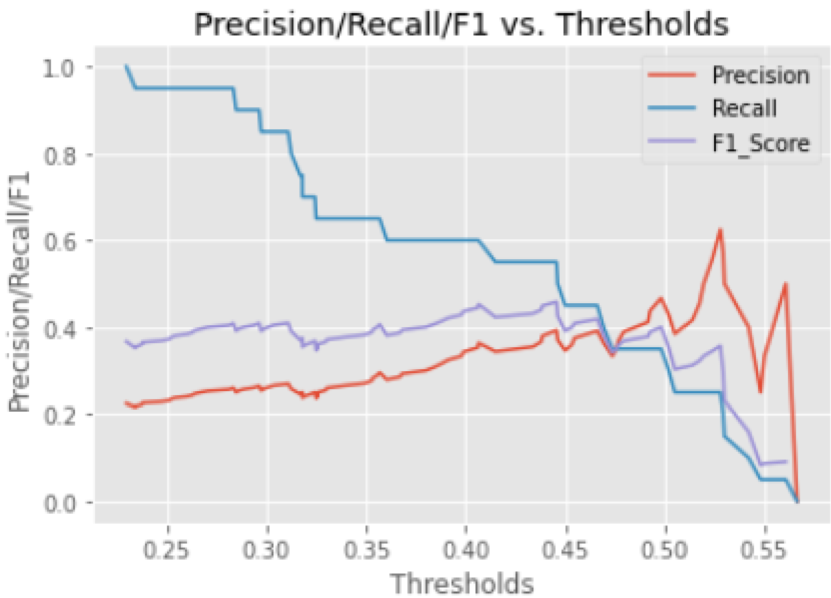
```
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1, activation='sigmoid'))
```



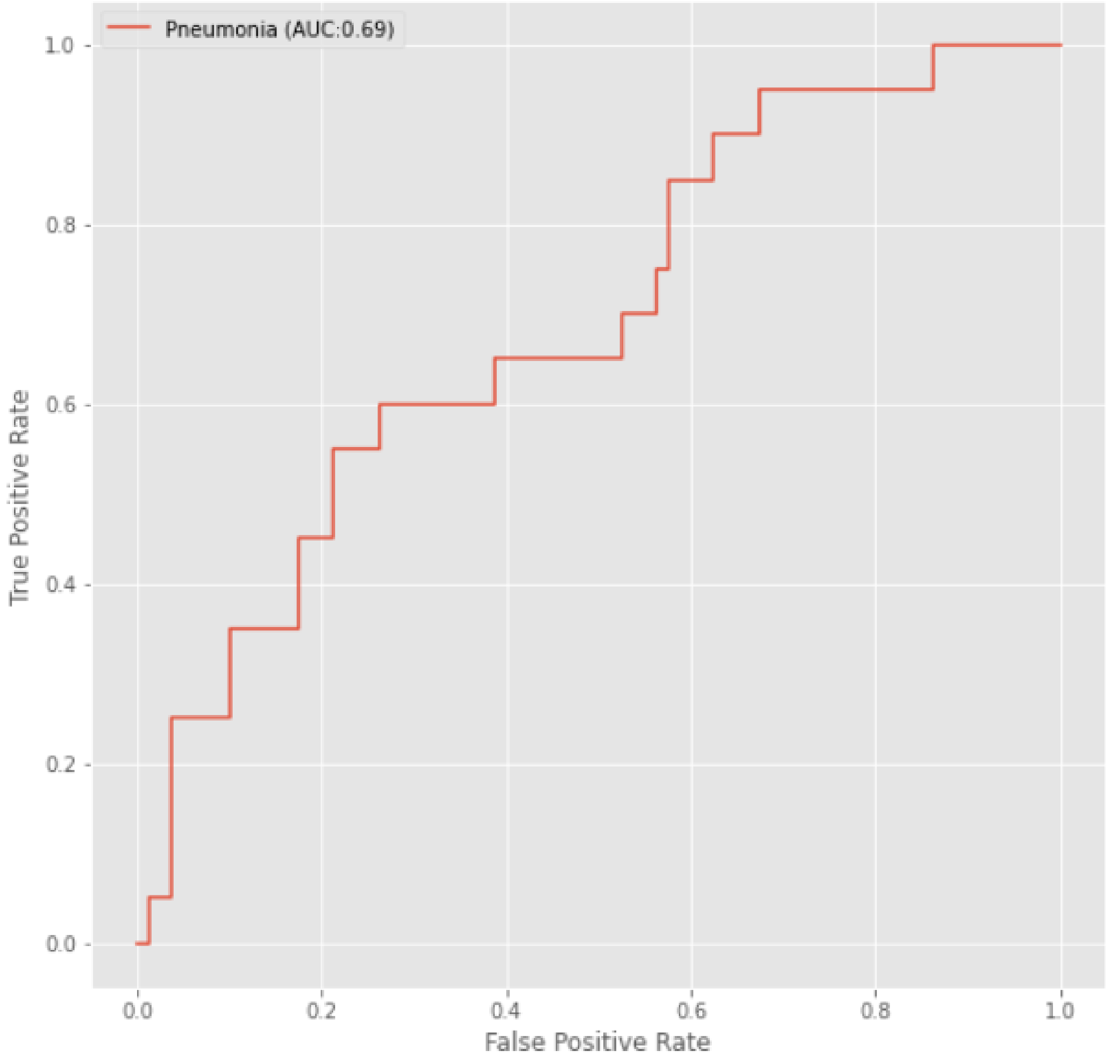
Final Threshold and Explanation:

The final threshold is 0.446 based on the maximum F1 for the chosen model with the intention of minimizing False Negatives.

```
Max F1 Score: 0.45833333333333337
Threshold corresponding to max F1: 0.4456912875175476
61
Threshold is: 0.4456913
```



```
fpr, tpr, thresholds_ROC, AUC =  plot_roc_curve(valY, pred_Y)
```



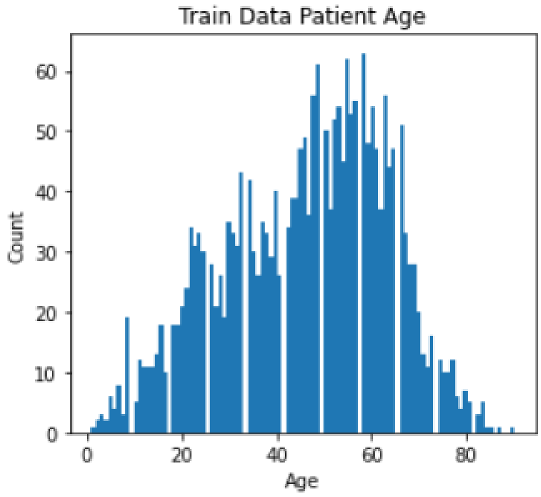
4. Databases

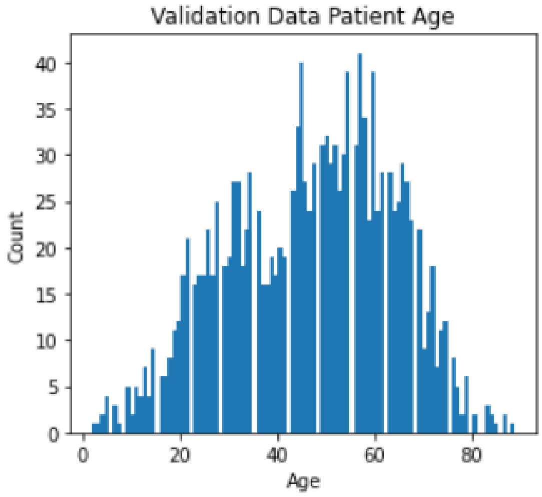
Description of Training Dataset:

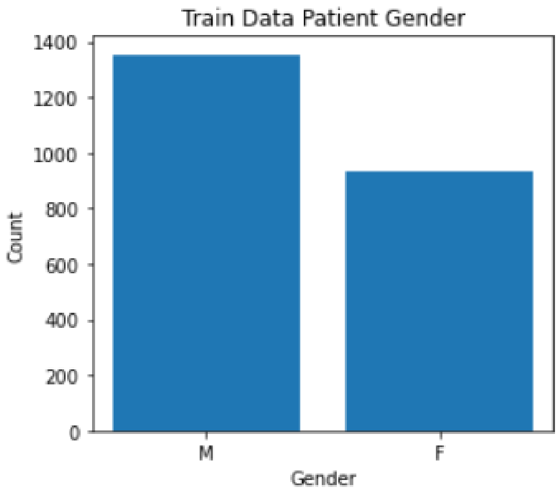
Train dataset has a total of 2288 records.
Train dataset has 1144 pneumonia cases which is 50.00% of the total training dataset.
Train dataset has 1144 no-pneumonia cases which is 50.00% of the total training dataset.

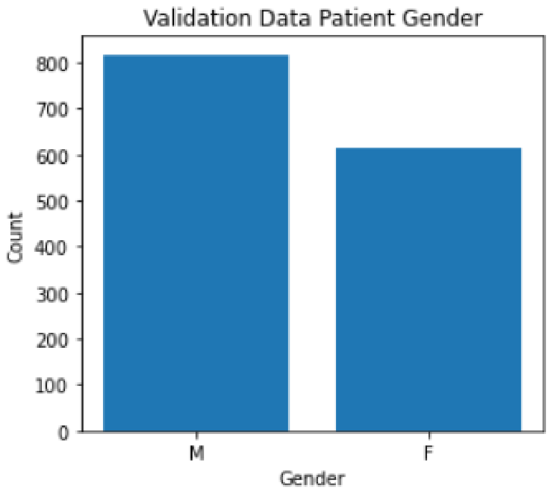
Description of Validation Dataset:

Validation dataset has a total of 1430 records.
Validation dataset has 286 pneumonia cases which is 20.00%.
Validation dataset has 1144 no-pneumonia cases which is 80.00%.









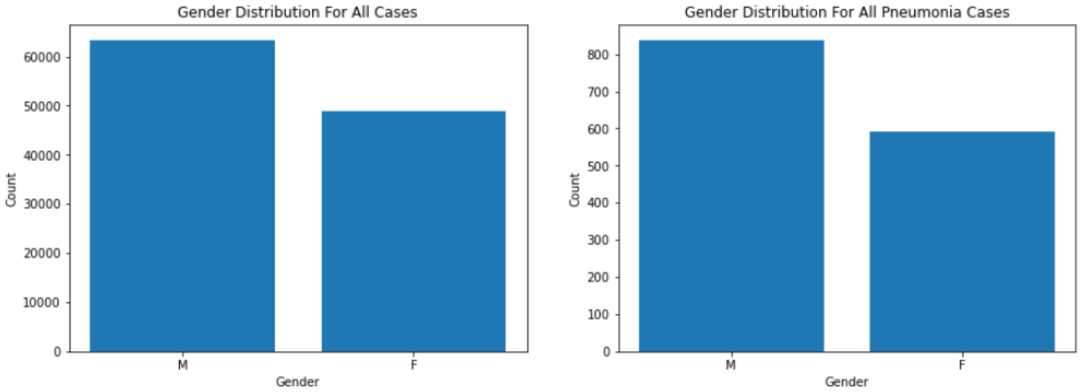
5. Ground Truth

The dataset has 112,120 X-ray images taken from 30,805 patients curated by the NIH. The disease labels were created using NLP mining the the associated radiological reports. The diseases include the common thoracic pathologies.

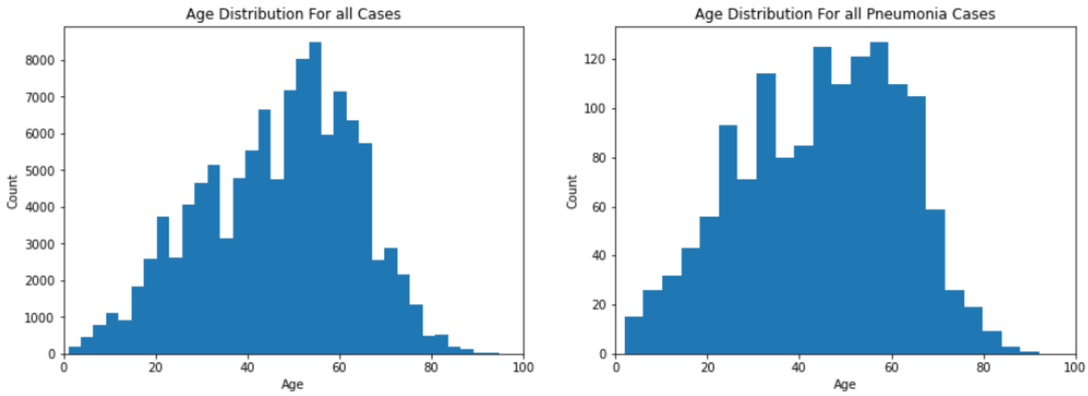
Mining the radiological reports to label the diseases using NLP was a major challenge for the algorithm. Thus erroneous labels were created, however, the accuracy was estimated over 90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

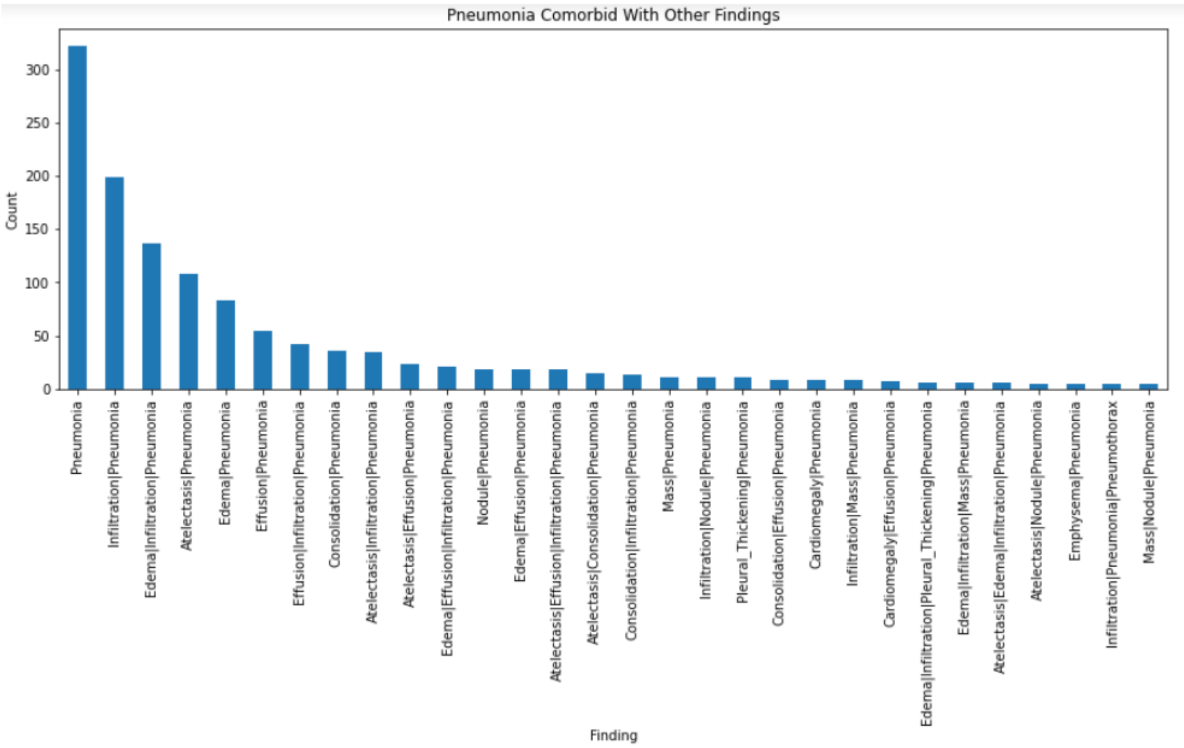


Seems like the ratios for all cases and pneumonia cases are about the same for gender. Males seem to account more for both population sets.



Most patients are in their 60's same with pneumonia patients.

Patients might have these comorbid diseases as depicted in the chart below.



Ground Truth Acquisition Methodology:

The Ground Truth would be a thorough evaluation of the X-Ray from several radiologists.

Algorithm Performance Standard:

Based on the published project, "[CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning]", the team was able to achieve an F1 score of 0.435 which can be considered as the benchmark performance.