

FDA Submission

****Your Name:****

Vinay Raj Singh Solanki

****Name of your Device:****

Pneumonia Detection Assistant for Radiologists from X-Ray Images

Algorithm Description

1. General Information

****Intended Use Statement:****

The algorithm assists the radiologist in detection of pneumonia from chest x-ray images.

****Indications for Use:****

The algorithm is intended to be used on male/female subjects between the age of 1-95 years on X-ray film of chest (PA and AP view) to help assist radiologists in predicting or detection of pneumonia.

****Device Limitations:****

It is not recommended for subjects/patients who also have Edema, Effusion or Infiltration, as these comorbidities have a similar pixel profile.

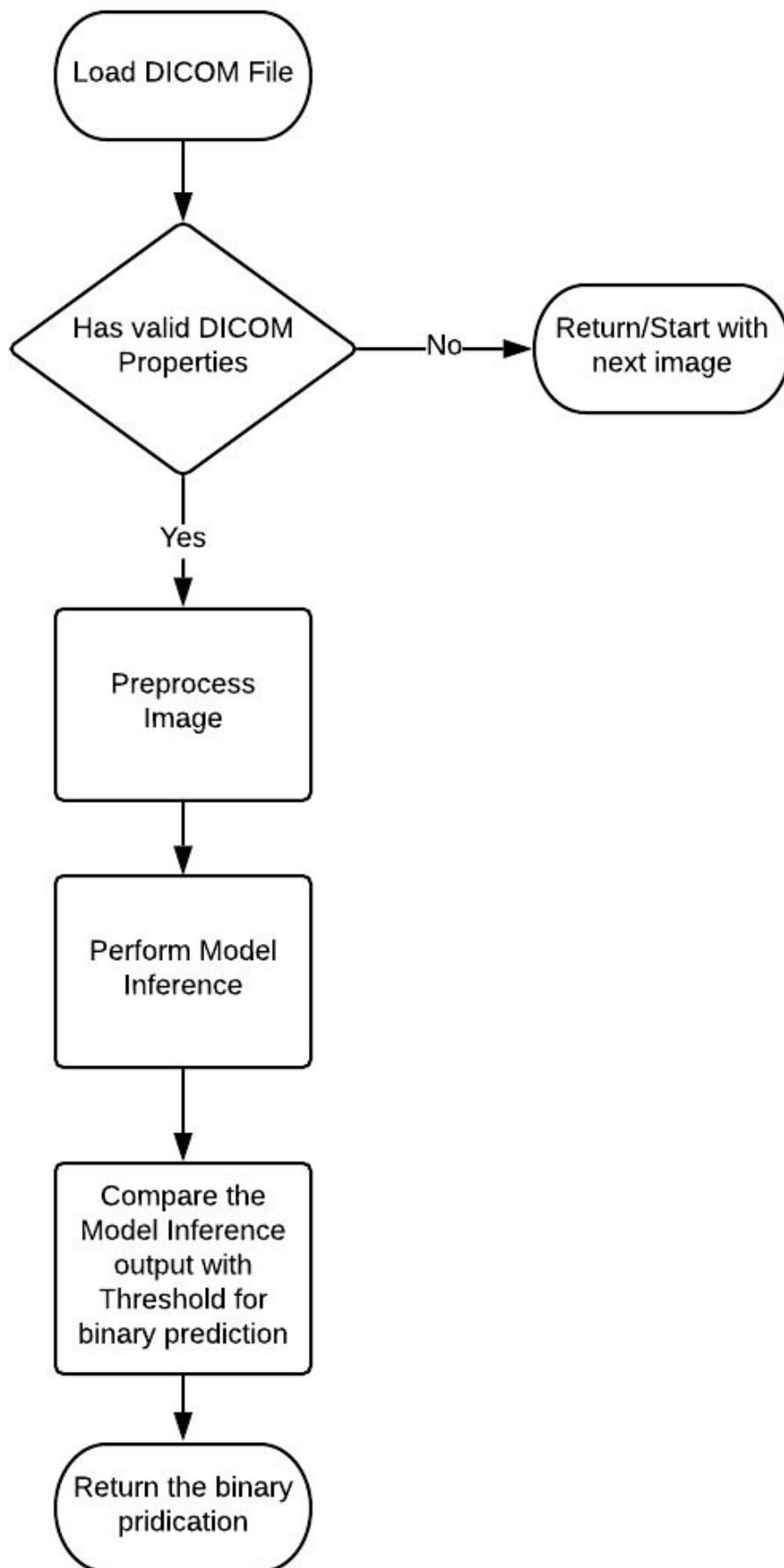
****Clinical Impact of Performance:****

With the selected algorithm, at threshold of 0.47, the precision is 0.36, recall is 0.75, Negative Prediction value is 0.92. We are able to correctly predict the 75 percent of positive cases, but also the False Positives reported are also very high. Also, the algo can identify 92

% (Negative Predicted value = 0.92) of negative cases correctly as True Negatives with few False negatives. So the model is very good prioritizing the records for Radiologist assistance.

2. Algorithm Design and Function

![Algorithm Flowchart](/algo_flowchart.png)



****DICOM Checking Steps:****

Confirm:

1. Patient is between ages 2-90
2. Patient is Male or Female
3. X-ray is of body part "chest"
4. Patient position is 'PA' (Posterior/Anterior) or 'AP' (Anterior/Posterior)
5. Modality is "DX" (Digital X-ray)

****Preprocessing Steps:****

Pre processing was done on both the Metadata and the Images:

**Metadata:*

1. Split the "Finding Labels" column and created one-hot-encoding, for each of the diseases independently
2. Removed all (around 16 records) records which has age greater than 95
3. Created a column "pneumonia_class" with values 1 and 0.
4. Split the Test and Validation set by ratio of 80:20
5. Balanced the "Pneumonia" and no pneumonia to be equal in the training data set (50:50).
6. Balanced the validation set to 20-80%, 20% positive and 80% negative.

**Images:*

1. Resized images to consistent square (224, 224) image size
2. Normalized the intensity (between 0 and 1 from original range of 0-255)
3. For training images, use horizontal flip and also shear, zoom, rotation height shift and width shift randomly by 10 percent.

****CNN Architecture:****

![[Algo Architecture]/vgg16_model_archi.png]

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359008
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359008
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359008
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359008
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359008
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
=====		
Total params: 14,714,688		
Trainable params: 4,719,616		
Non-trainable params: 9,995,072		
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, 1)	257
=====		
Total params: 41,062,209		
Trainable params: 31,067,137		
Non-trainable params: 9,995,072		

3. Algorithm Training

****Parameters:****

** Types of augmentation used during training*

1. rescale pixels from 0 to 1, by dividing /255
2. horizontal flip
3. height shift range: 0.1
4. width shift range: 0.1
5. rotation range: 10
6. shear range: 0.1
7. zoom range: 0.1

** Types of augmentation used for validation data:*

1. rescale pixels from 0 to 1, by dividing /255

** Batch size: 64*

** Optimizer learning rate: Adam Optimizer with learning rate of 1e-6*

** Layers of pre-existing architecture that were frozen:*

1. block1_conv1 False
2. block1_conv2 False
3. block1_pool False
4. block2_conv1 False
5. block2_conv2 False
6. block2_pool False
7. block3_conv1 False
8. block3_conv2 False
9. block3_conv3 False
10. block3_pool False
11. block4_conv1 False
12. block4_conv2 False
13. block4_conv3 False
14. block4_pool False
15. block5_conv1 False

** Layers of pre-existing architecture that were fine-tuned:*

1. block5_conv2 True
2. block5_conv3 True
3. block5_pool True

** Layers added to pre-existing architecture:*

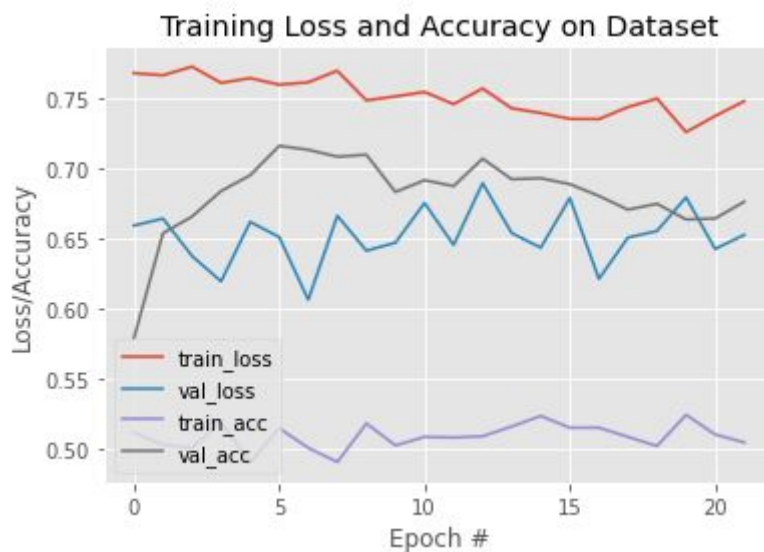
1. Flatten
2. Dropout(0.45)
3. Dense layer, 1024 nodes, relu activation
4. Dropout(0.4)
5. Dense layer, 512 nodes, relu activation
6. Dropout(0.3)
7. Dense layer, 256 nodes, relu activation
8. Dropout(0.2)

9. Dense layer, 1 node, sigmoid

Training history: the model ran for 22 epochs, model on epoch 7 was saved as best among these.

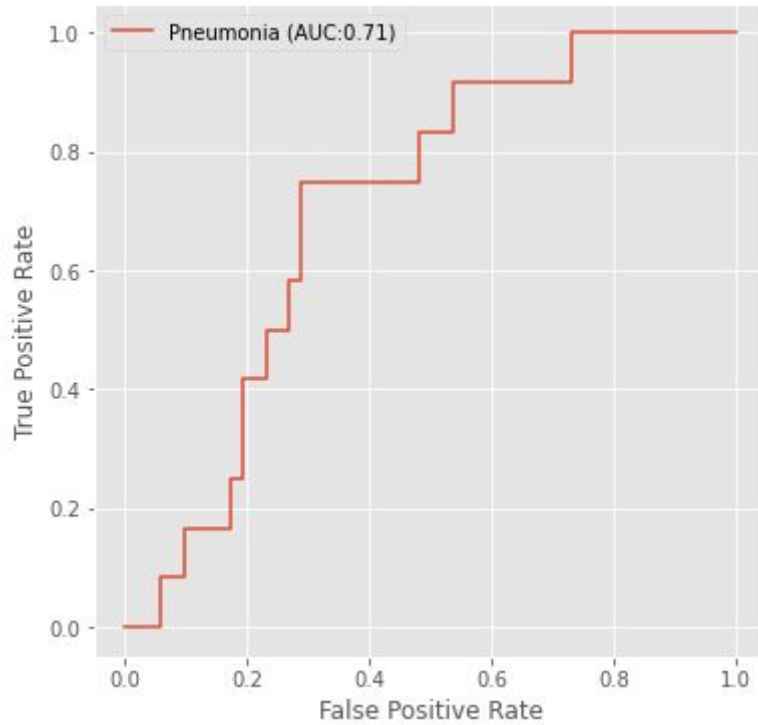
1. Training loss, training accuracy, validation loss and validation accuracy while training the model

![[Training loss, training accuracy, validation loss and validation accuracy while training the model]](/training_loss_accuracy.png)



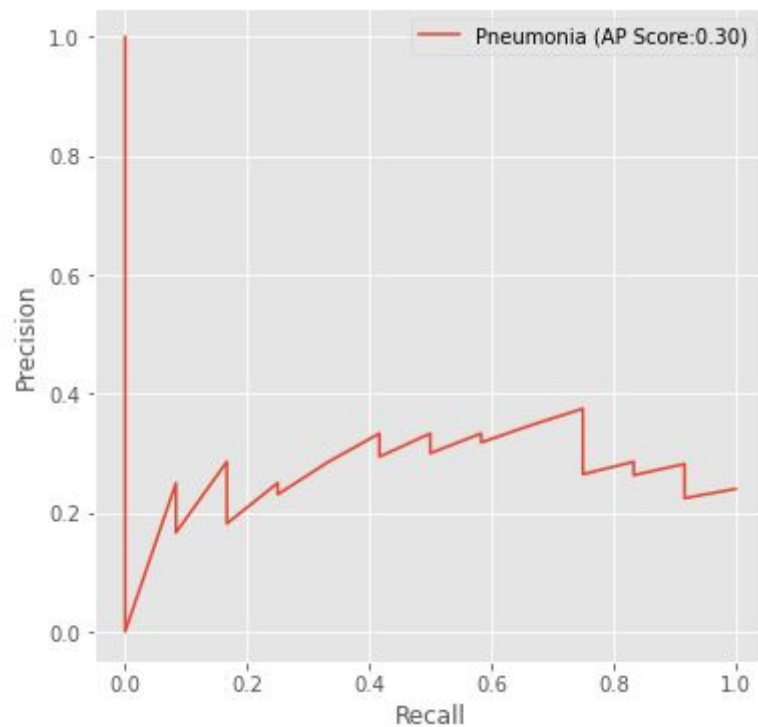
2. True Positive Rate vs False Positive Rate of model

![[True Positive Rate vs False Positive Rate of model]](/tpr_vs_fpr_curve.png)



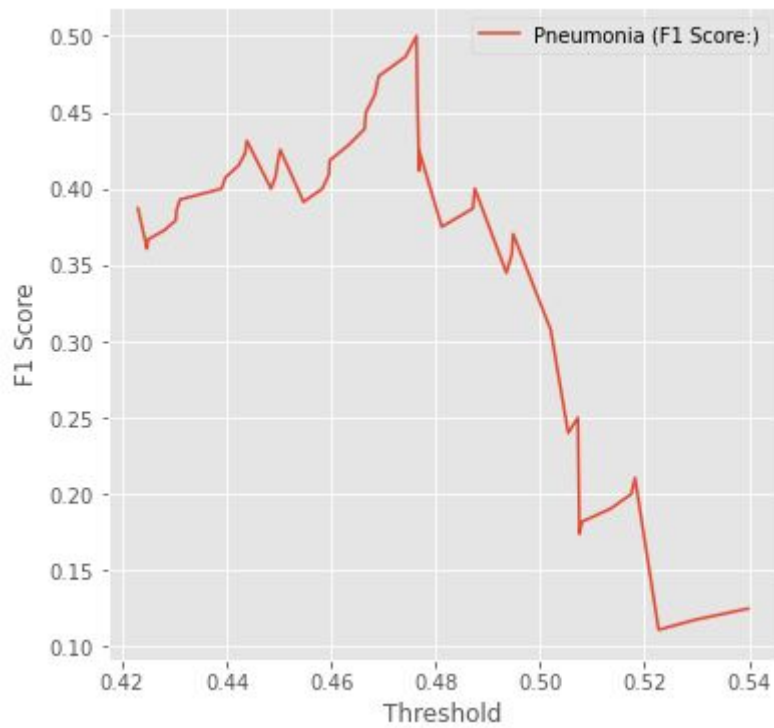
3. Precision vs Recall of model

![[Precision vs Recall of model]](/precision_vs_recall.png)



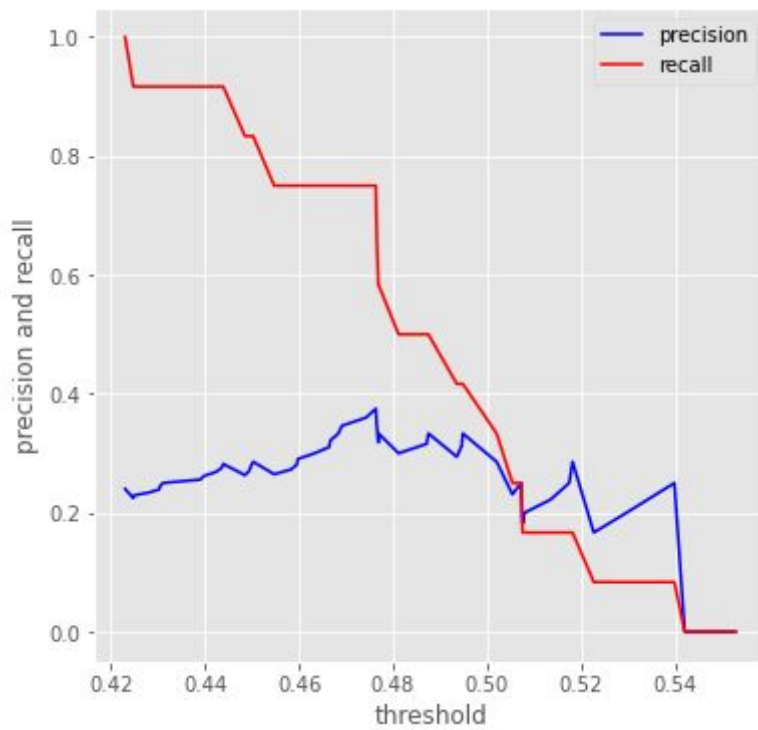
4. f1-score vs threshold of model

![[f1-score vs threshold of model]](/f1score_vs_threshold.png)



5. precision and recall vs threshold of model

![precision and recall vs threshold of model](/precision_and_recall_vs_threshold.png)



****Final Threshold and Explanation:****

The final threshold was determined by finding that value that maximized F1 score, which is also known as the harmonic mean of precision and sensitivity. So, the final threshold achieves a balance between the largely competing interests of these two component measures.

Final Threshold we chose = 0.47, with this threshold, the recall is 0.75 and precision is 0.36. The confusion matrix with this threshold is: Threshold:0.47 :- True Positive: 9, False Positive: 16, False Negative: 3, True Negative: 36

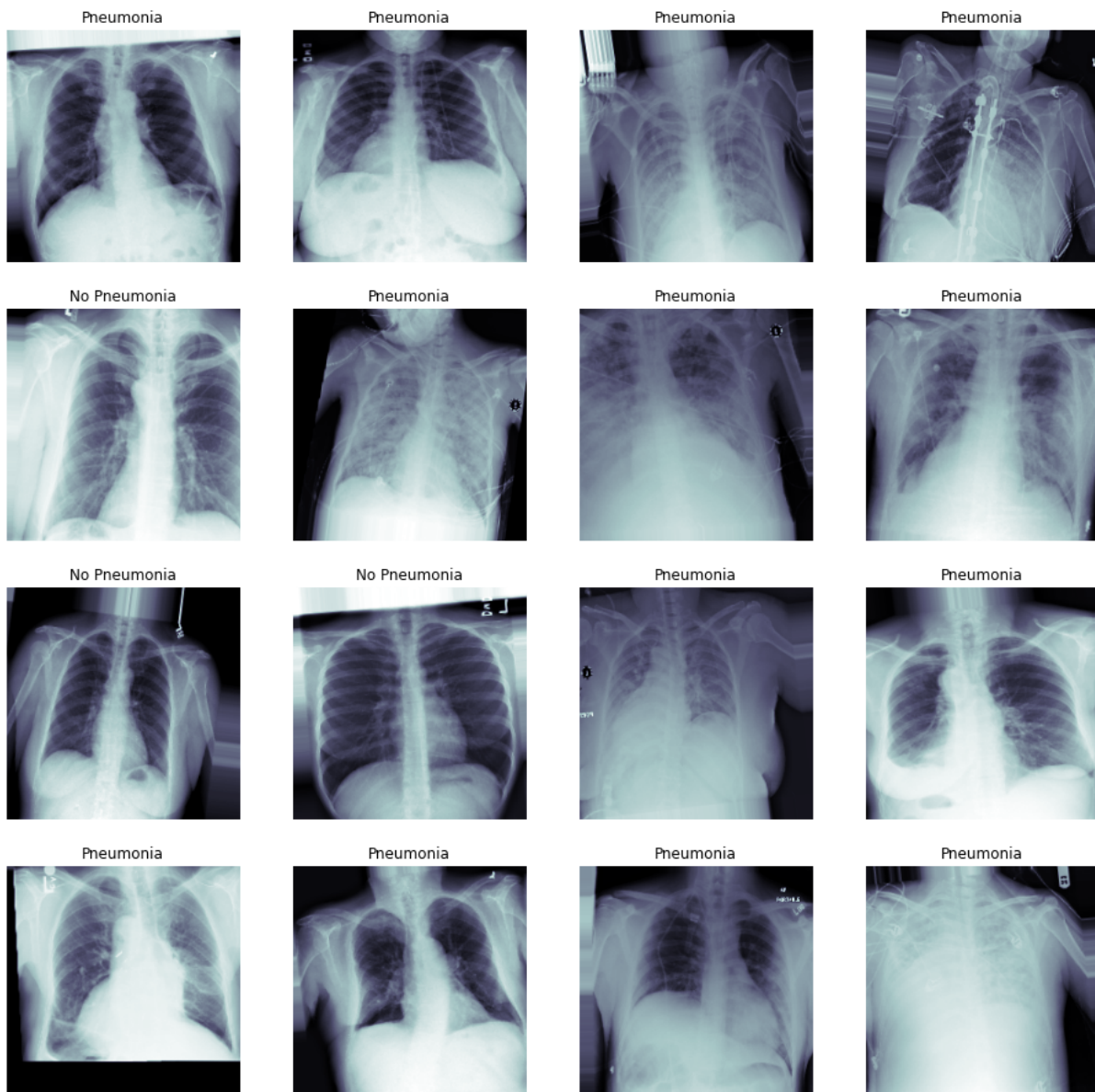
4. Databases

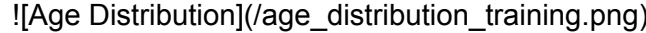
(For the below, include visualizations as they are useful and relevant)

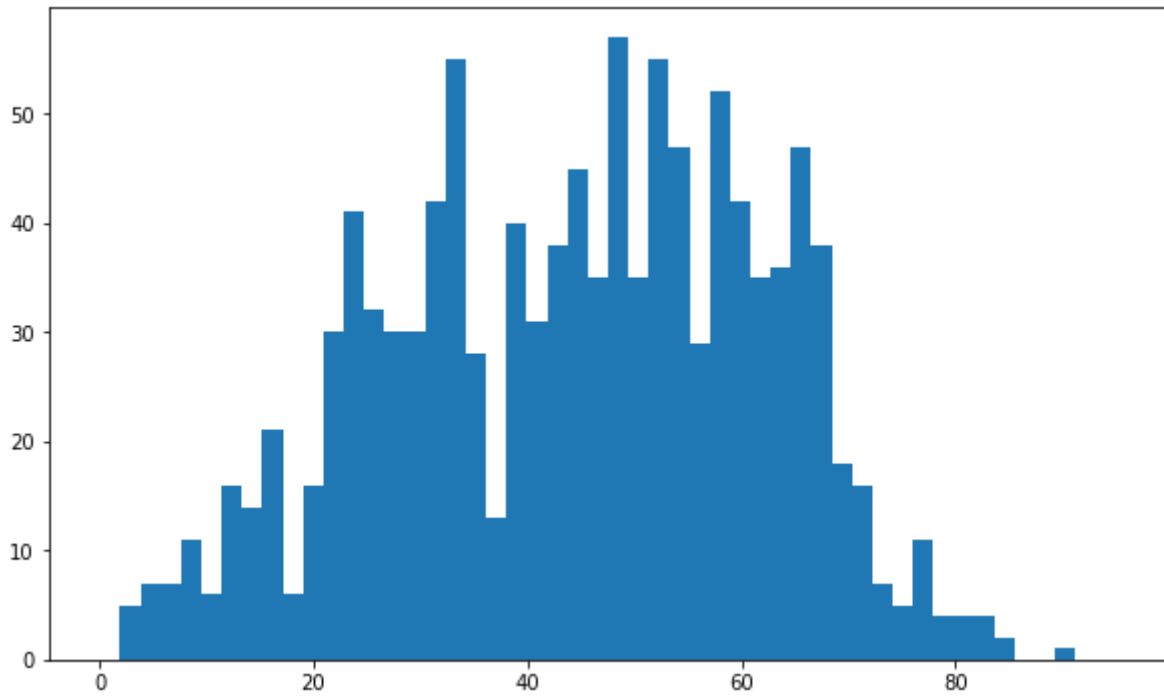
****Description of Training Dataset:****

1. Training dataset consisted of 2288 chest X-ray images, with a 50/50 split between positive and negative cases.

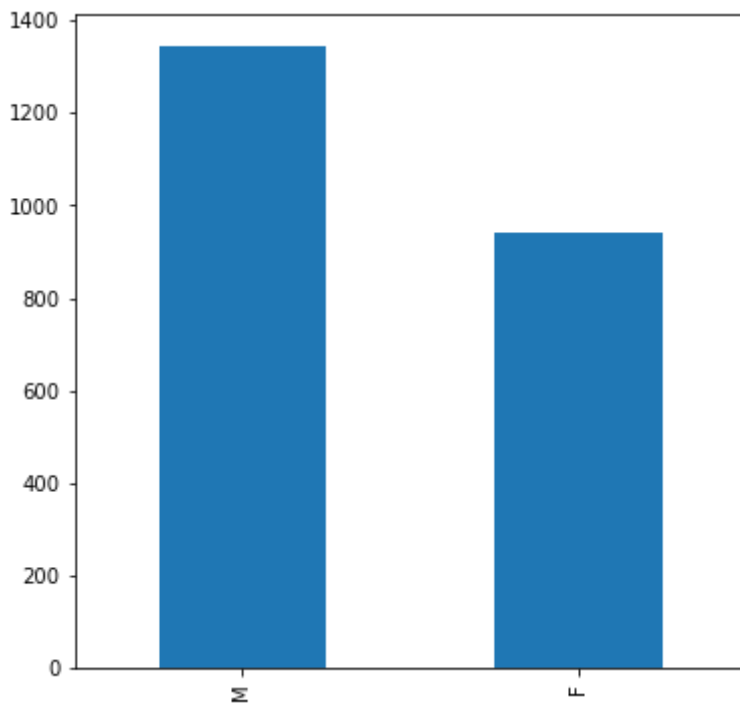
![Sample Images](/samples_images.png)



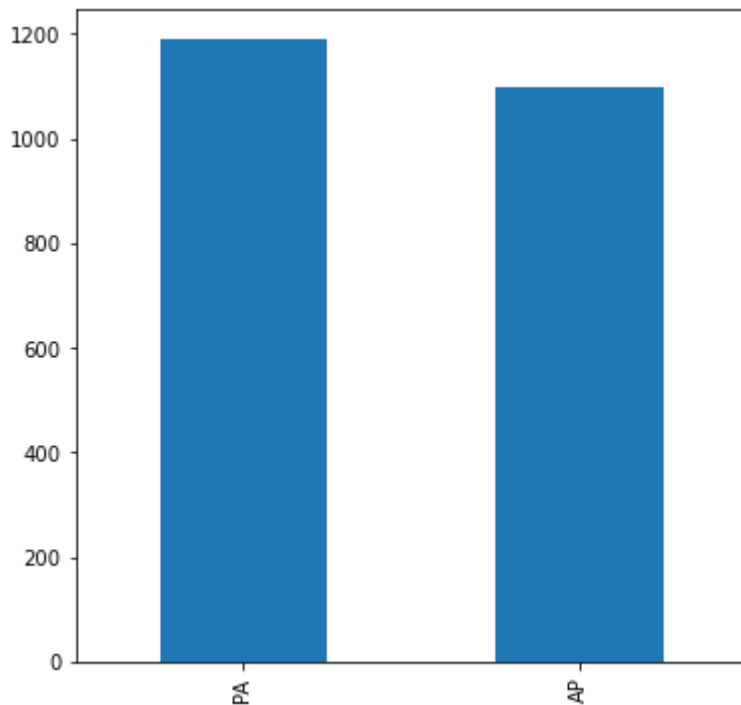
2. Patient ages ranged between 1-95 years old.




3. 58.82% of patients were male and 41.17% were female.
 ![Gender Distribution](/gender_distribution_training.png)

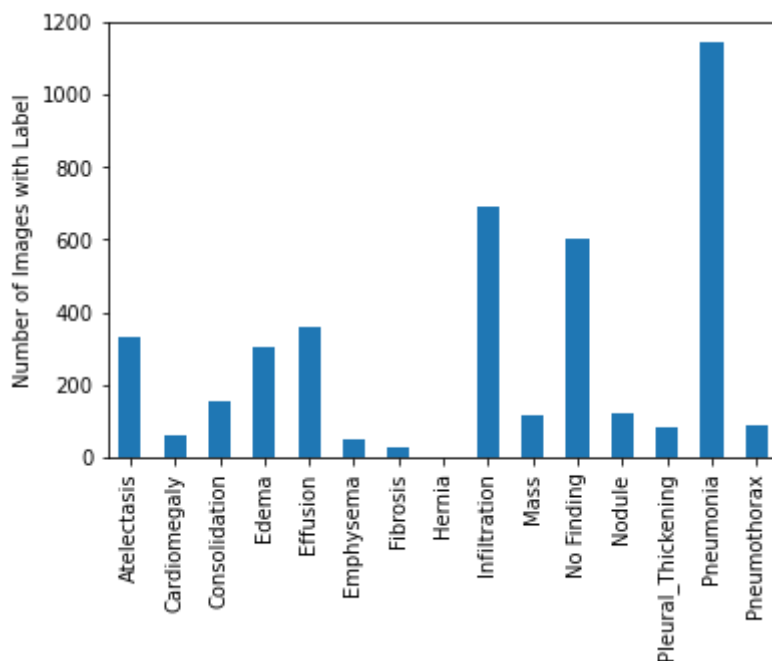


4. The images were taken in PA (51.96%) and AP(48.03%) positions.
 ![Position Distribution](/position_distribution_training.png)



5. In the population, the three most prevalent conditions were Infiltration, Effusion and Atelectasis. Among patients with Pneumonia, the most prevalent comorbidities were Infiltration, Edema, Effusion and Atelectasis. Pneumonia diagnostic was slightly correlated with Edema.

![Position Distribution](/desease_distribution_training.png)

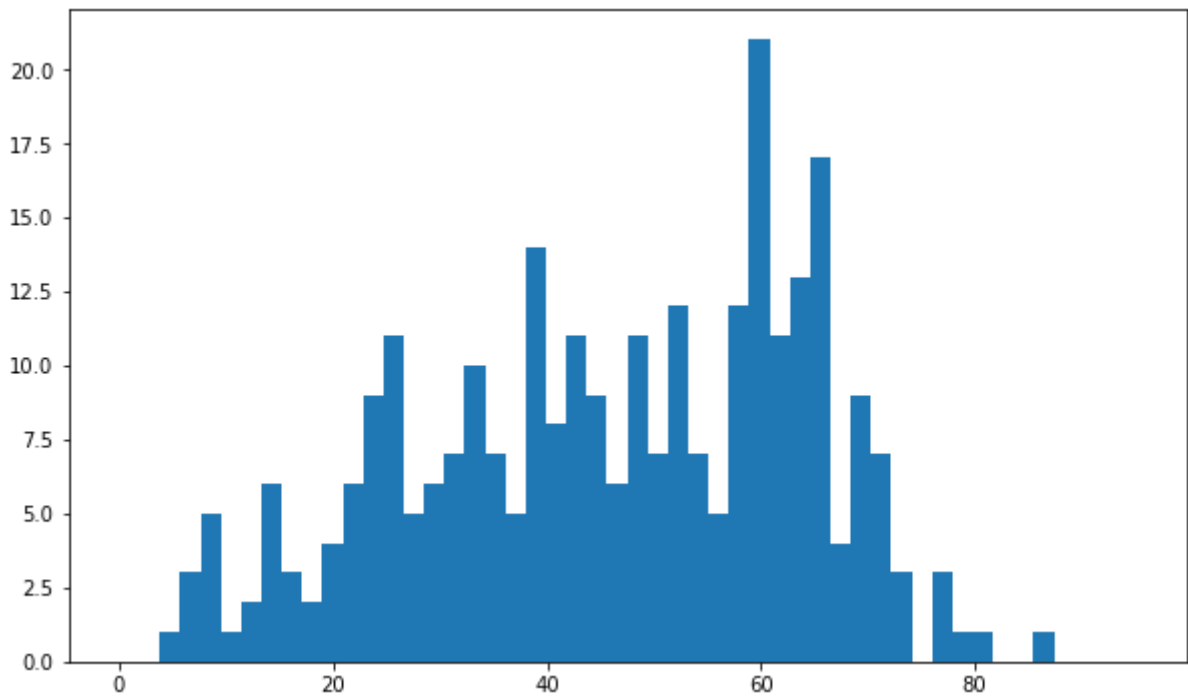


****Description of Validation Dataset:****

1. Validation dataset consisted of 1430 chest X-ray images, with a 20/80 split between positive and negative cases to reflect the actual prevalence of pneumonia.

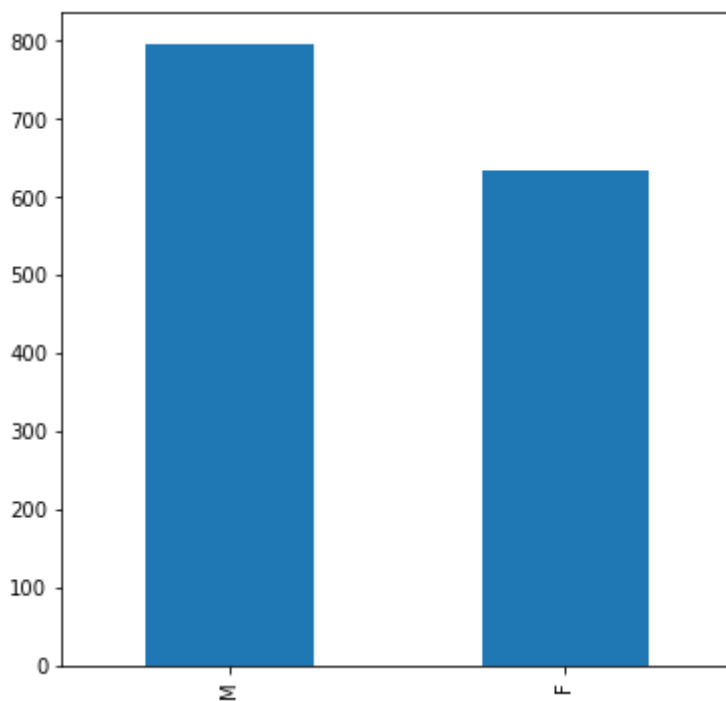
2. Patient ages ranged between 3-93 years old.

![Age Distribution](/age_distribution_valid.png)



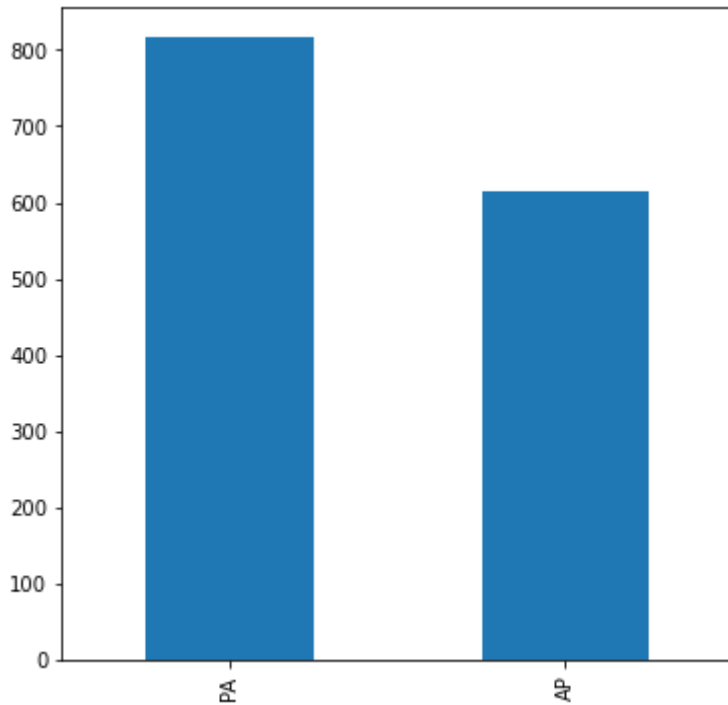
3. 55.73% of patients were male and 44.26% were female.

![Gender Distribution](/gender_distribution_valid.png)



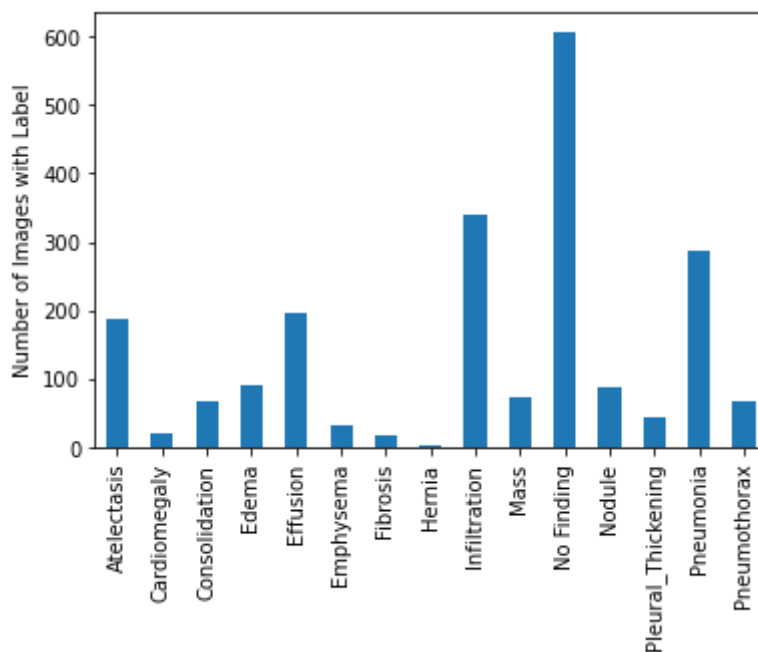
4. The images were taken in PA (57.06%) and AP(42.93%) positions.

![Position Distribution](/position_distribution_valid.png)



- In the population, the three most prevalent conditions were Infiltration, Effusion and Atelectasis. Among patients with Pneumonia, the most prevalent comorbidities were Infiltration, Edema, Effusion and Atelectasis. Pneumonia diagnostic was slightly correlated with Edema.

![Position Distribution](/disease_distribution_valid.png)



5. Ground Truth

From the Kaggle data source (<https://www.kaggle.com/nih-chest-xrays/data>):

This NIH Chest X-ray Dataset comprises 112,120 X-ray images with disease labels from 30,805 unique patients.

To create these labels, 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.

This means the labels are mostly accurate for training. However, a more accurate alternative would be to have 3 radiologists verify the diagnoses and label the images, possibly weighted by years of experience. This would be a very lengthy and costly procedure for 112,120 images. Given the training required, one cannot simply crowdsource the labels like one can for common objects such as those found in ImageNet. Thus, if the NLP approach is 90% accurate, it is likely a good compromise.

Disease labels include: atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening, cardiomegaly, nodule, mass, and hernia.

6. FDA Validation Plan

****Patient Population Description for FDA Validation Dataset:****

Patient population to be :

1. Age range: 1-95 years old
2. Sex to be male or female (approximately 58:42 ratio for Male:Female)
3. Type of imaging Digital X-rays
4. Body part image (X-ray) taken: Chest in PA/AP positions
5. Patients with effusion, consolidation or edema may not be suitable for ruling out pneumonia
6. The validation data set should contain a 20% prevalence of pneumonia.

****Ground Truth Acquisition Methodology:****

So Recruiting 7 certified radiologists to identify pneumonia or 'no pneumonia' labels for the entire image data set. It's expected that a minimum of 5 would provide labels for any particular image, so that 3 or more would constitute a 'label' decision.

****Algorithm Performance Standard:****

Algorithm's performance can be measured by calculating F1 score against 'silver standard' ground truth as described above. The algorithm's F1 score should exceed 0.387 (which is average of 4 radiologist given in the CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays With Deep Learning, paper at <https://arxiv.org/pdf/1711.05225.pdf>), where a similar method is used to compare device's F1 score to average F1 score over four radiologists.