FDA Submission

Your Name: Ahmed M. Fathy
Name of your Device: Pneumonia Detection from Chest X-Rays software.

Algorithm Description

1. General Information

Intended Use Statement:

- This algorithm is intended for use on both women and men from the ages of 5-85 who have been administered a chest X-Rays pneumonia study using PA or AP position on a X-Rays machine.

Indications for Use:

- for assisting radiologists in the **conferming** of pneumonia on chest X-Rays images.

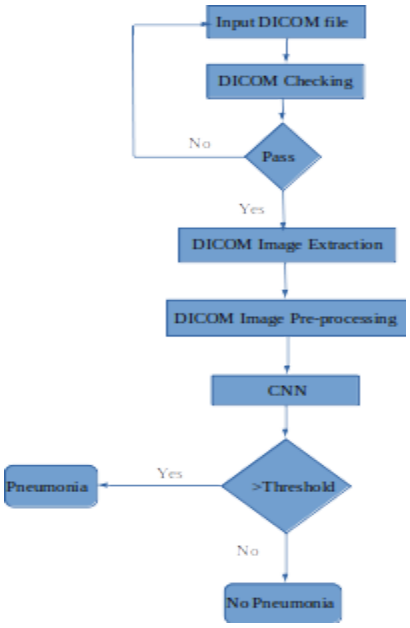
Device Limitations:

- The Algorithm has a high false negative rate, so it is not recommended for using in screening or prioritization.

Clinical Impact of Performance:

- This algorithm has a high precision (low false positive rate), so it can be used for helping radiologists in **confirming** pneumonia diagnosis.

2. Algorithm Design and Function



DICOM Checking Steps:

- Patient's age between 5 and 85.
- Patient's examined body part is chest.
- Patient's position is PA or AP

Preprocessing Steps:

- Images are scaled by 1/255
- Images are resized to [224,244,3]

CNN Architecture:

Layer type	Output Shape	Param	note
InputLayer	(None, 224, 224, 3)	0	VGG16(freezed)
Conv2D	(None, 224, 224, 64)	1792	VGG16(freezed)
Conv2D	(None, 224, 224, 64)	36928	VGG16(freezed)
MaxPooling2D	(None, 112, 112, 64)	0	VGG16(freezed)
Conv2D	(None, 112, 112, 128)	73856	VGG16(freezed)
Conv2D	(None, 112, 112, 128)	147584	VGG16(freezed)
MaxPooling2D	(None, 56, 56, 128)	0	VGG16(freezed)
Conv2D	(None, 56, 56, 256)	295168	VGG16(freezed)
Conv2D	(None, 56, 56, 256)	590080	VGG16(freezed)
Conv2D	(None, 56, 56, 256)	590080	VGG16(freezed)
MaxPooling2D	(None, 28, 28, 256)	0	VGG16(freezed)
Conv2D	(None, 28, 28, 512)	1180160	VGG16(freezed)
Conv2D	(None, 28, 28, 512)	2359808	VGG16(freezed)
Conv2D	(None, 28, 28, 512)	2359808	VGG16(freezed)
MaxPooling2D	(None, 14, 14, 512)	0	VGG16(freezed)
Conv2D	(None, 14, 14, 512)	2359808	VGG16(Tunned)
Conv2D	(None, 14, 14, 512)	2359808	VGG16(Tunned)
Conv2D	(None, 14, 14, 512)	2359808	VGG16(Tunned)
MaxPooling2D	(None, 7, 7, 512)	0	added

Flatten	(None, 25088)	0	added
Dense	(None, 1024)	25691136	added
Dropout	(None, 1024)	0	added
Dense	(None, 512)	524800	added
Dropout	(None, 512)	0	added
Dense	(None, 2)	1026	added

3. Algorithm Training

Parameters:

- Adam optimizer was used with learning rate 0.0001 and decay 1e-5.
- Binary cross entropy was used for the loss.
- Accuracy matrix was used.
- Number of epochs 15.

Types of augmentation used during training:

- rescale=1. / 255.0
- horizontal_flip
- Height_shift_range = 0.1
- Width_shift_range = 0.1
- Rotation_range = 20.0
- Shear_range = 0.1
- Zoom_range = 0.1

Batch size:

- Training = 16
- Validation = 32

Optimizer learning rate:

- 0.0001

Layers of pre-existing architecture that were frozen:

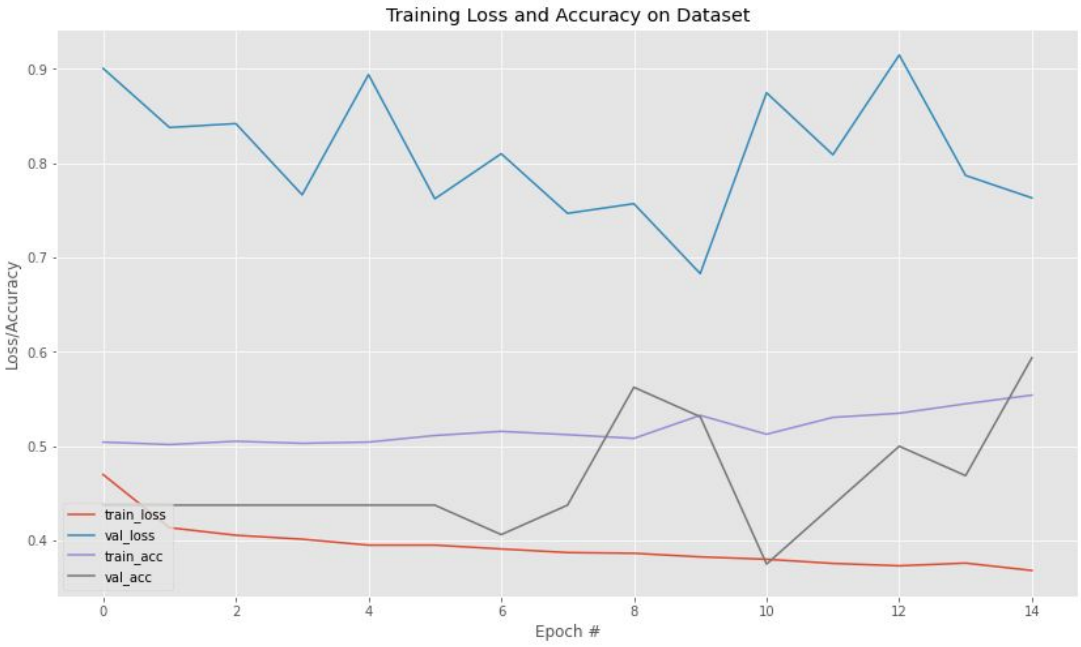
- The first 15 layers In VGG16 model (see the CNN Architecture).

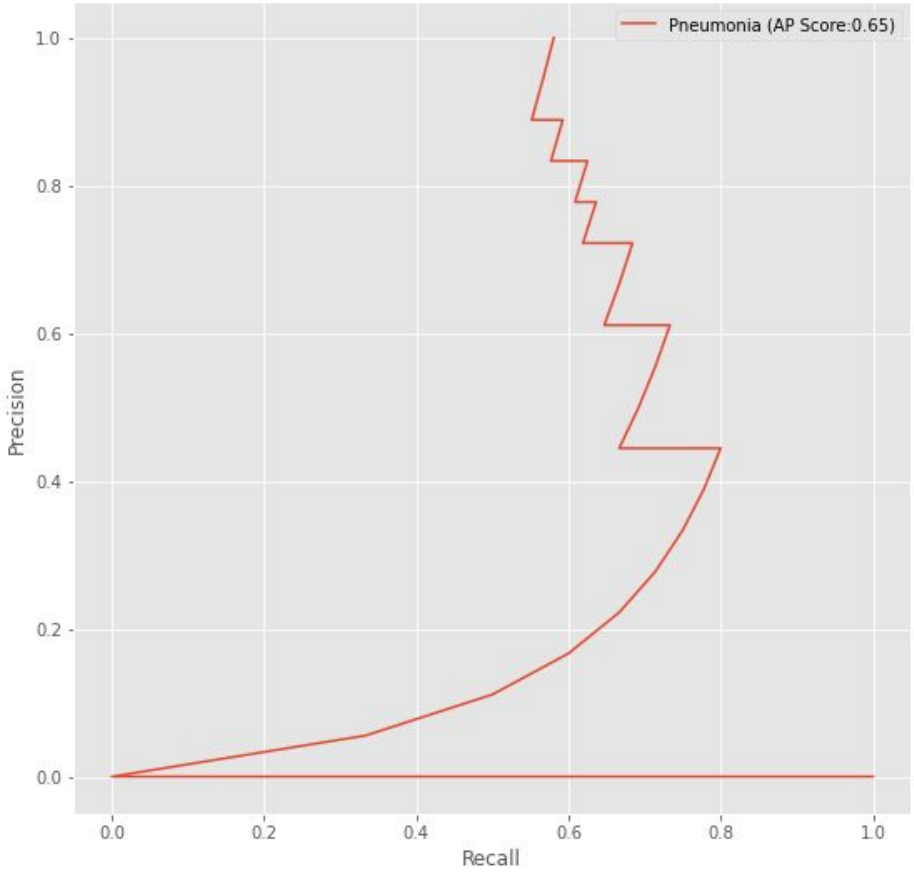
Layers of pre-existing architecture that were fine-tuned:

- The 3 convolution layers after 15 layer and the max pooling layer (see the CNN Architecture).

Layers added to pre-existing architecture:

- Flatten - Dense - Dropout - Dense - Dropout - Dense (see the CNN Architecture).

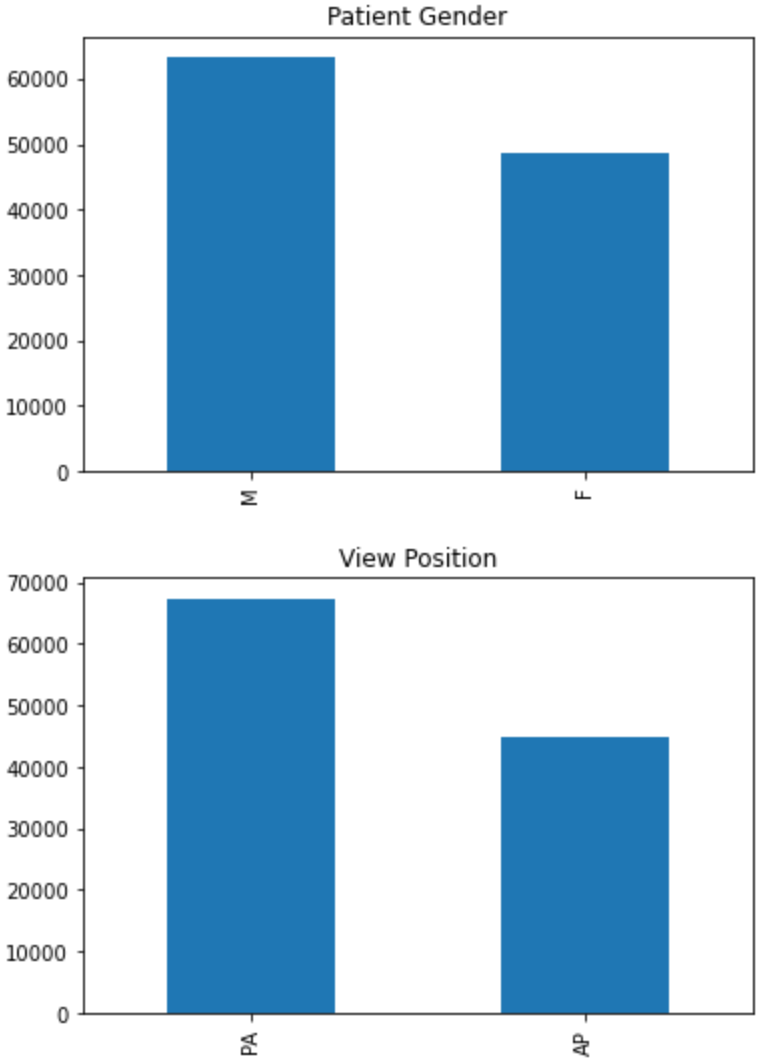


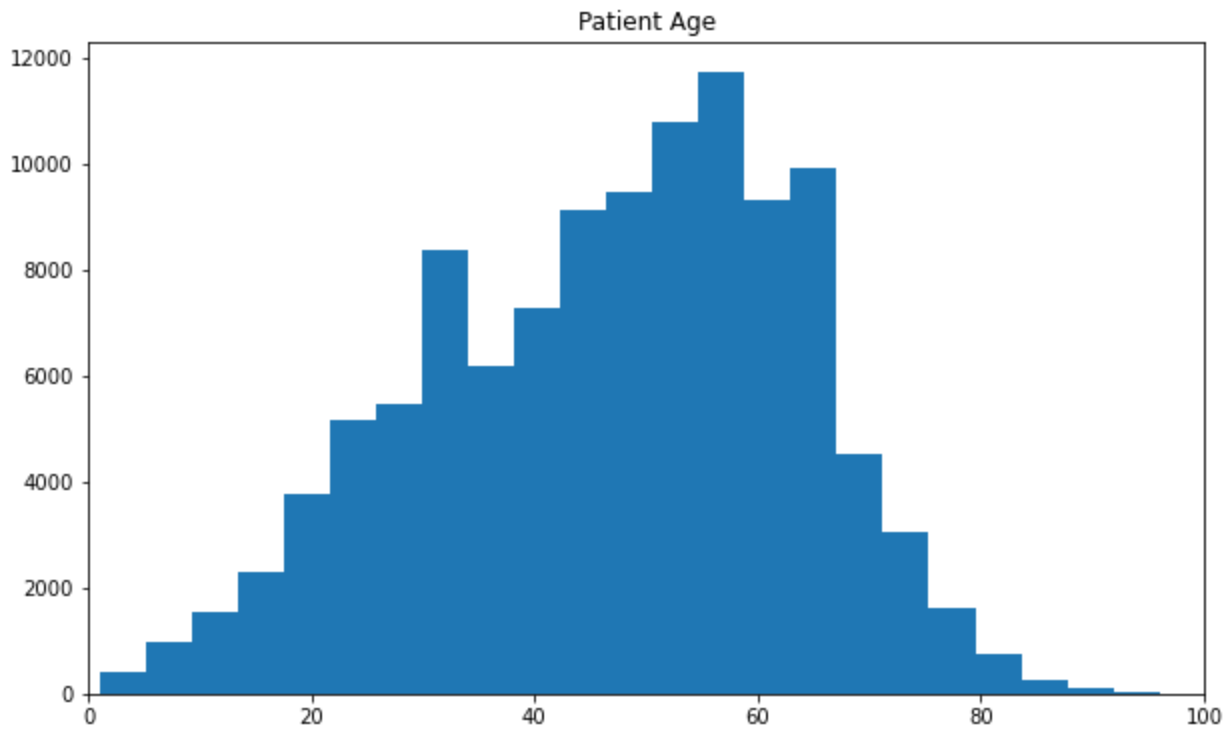


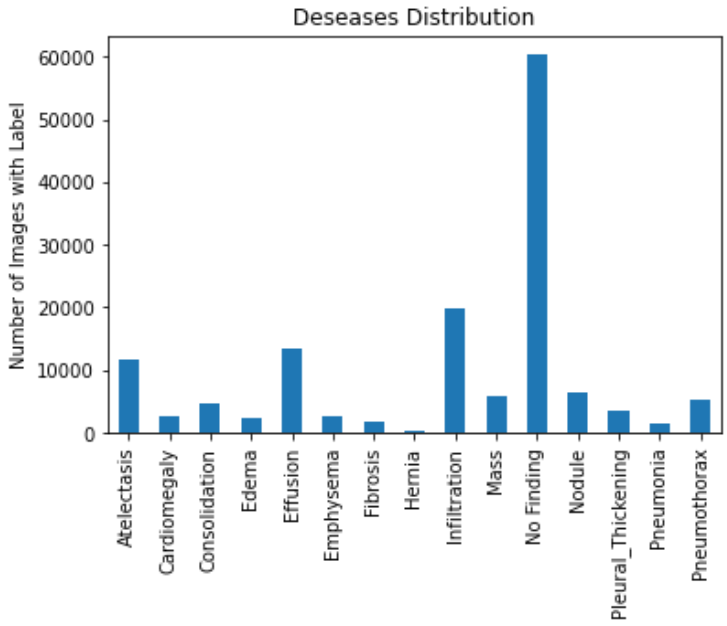
Threshold and Explanation:

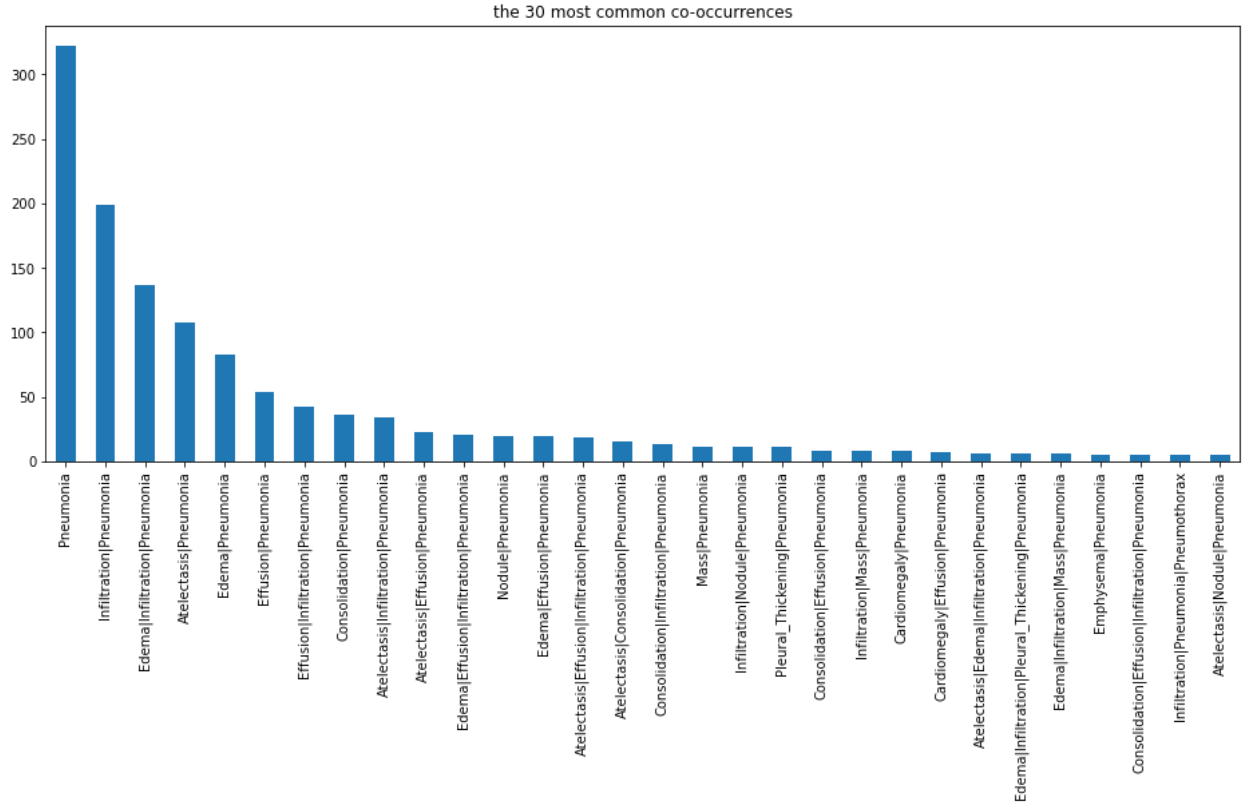
The threshold was chosen to be **0.51.** This achieves **0.8 precision** and **0.44 recall** and makes the algorithm efficient in **confirming** the pneumonia cases.

4. Databases


Patient Gender
60000
50000
40000
30000
20000
10000
0
M
F
View Position
70000
60000
50000
40000
30000
20000
10000
0
PA
AP


Patient Age
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100





Description of Training Dataset:

- The training dataset is 80% of the total dataset. It was balanced and has 2290 cases.

Description of Validation Dataset:

- The validation dataset is 20% of the total dataset. It was balanced and has 572 cases.

5. Ground Truth

- The disease labels were created using Natural Language Processing (NLP) to mine the associated radiological reports.
- The biggest limitation of this dataset is that image labels were NLP-extracted so there could be some erroneous labels but the NLP labeling accuracy is estimated to be >90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

- To validate the algorithms, I would collect a validation set that was made up of 2d X-rays chest images with both PA and AP positions for both women and men between the ages of 5 and 85. I would also want to make sure that the distribution of pneumonia in my validation set was reflective of the distribution of the density that is seen in the real world.

Ground Truth Acquisition Methodology:

- As the diagnosis of pneumonia from chest X-rays is difficult, **the silver standard approach** of using several radiologists was used.

Algorithm Performance Standard:

- The algorithm F1 score is 0.57, which is equal to the mean of F1 scores published in this work.