FDA Submission

Your Name: Julie Zhang

Name of your Device: PneumoniaDetectionHelper

Algorithm Description

1. General Information

Intended Use Statement: Assist the radiologist in the detection of Pneumonia on chest x-rays.

Indications for Use: Screening Pneumonia studies for woman and man between the ages of 10-80 years old with no prior history of Pneumonia

Device Limitations: Need GPU for computation

Clinical Impact of Performance:

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

The algorithm is designed for screening studies and we are interested to reduce False Negative. Therefore, we set our threshold in favor of maximizing the recall.

2. Algorithm Design and Function
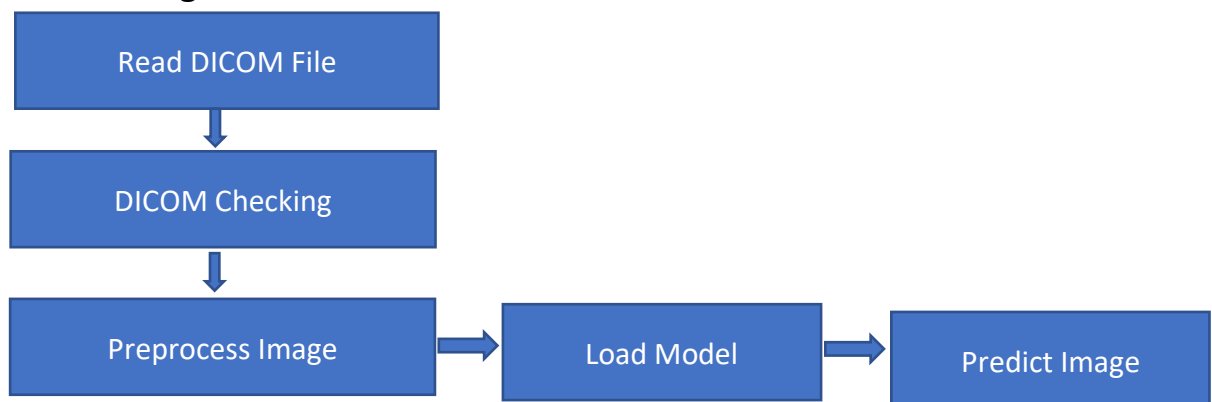
DICOM Checking Steps:
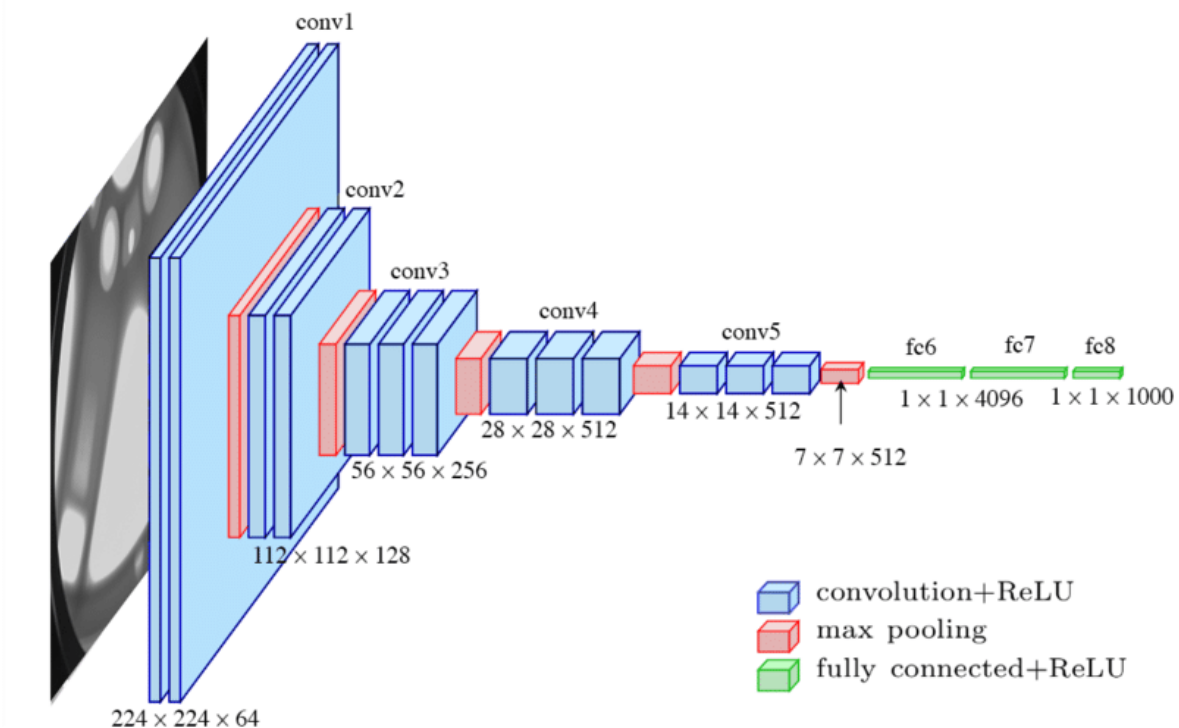The algorithm performs the following checks on the DICOM image:

Check Patient Age is between 10 and 80.

Preprocessing Steps:
Re-sizes the image to 244 x 244
standardize the intensity to be between 0 and 1 (from original range of 0 to 255)

CNN Architecture:

The built CNN uses the VGG model in the above image till the last max pooling layer and adds a fatten layer and a dense layer.

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

- o width_shift_range=0.1
- o rotation_range=20
- o shear_range=0.1
- o zoom_range=0.1

- **Batch size:** 32
- **Optimizer learning rate:** Adam(lr=1e-4)
- **Layers of pre-existing architecture that were frozen:**

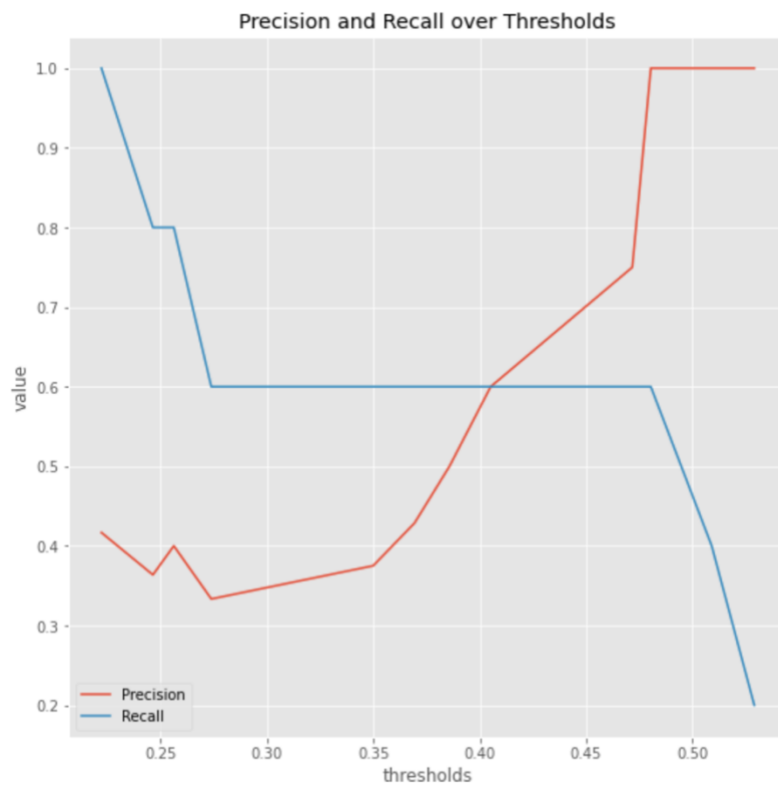| Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 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)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0

- **Layers added to pre-existing architecture:**

```
_____
flatten_1 (Flatten)             (None, 25088)              0
_____
dense_1 (Dense)                 (None, 1)                 25089
=================================================================
```

- **Algorithm training performance visualization**

Training Loss and Accuracy on Dataset
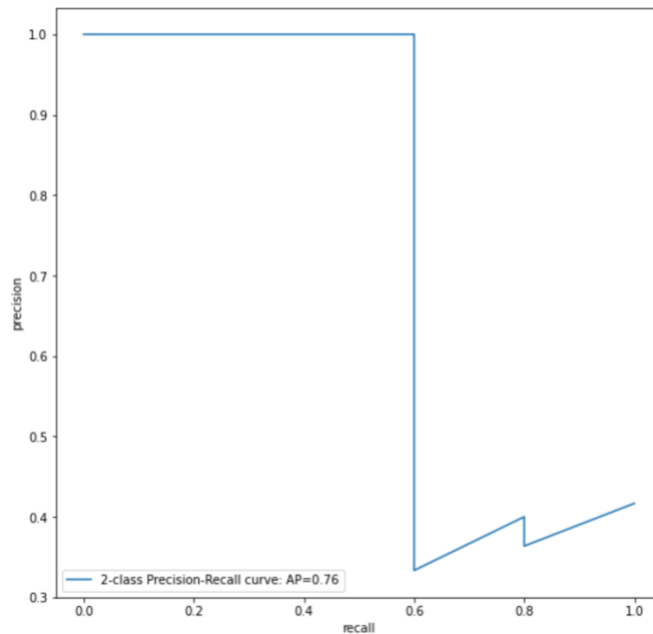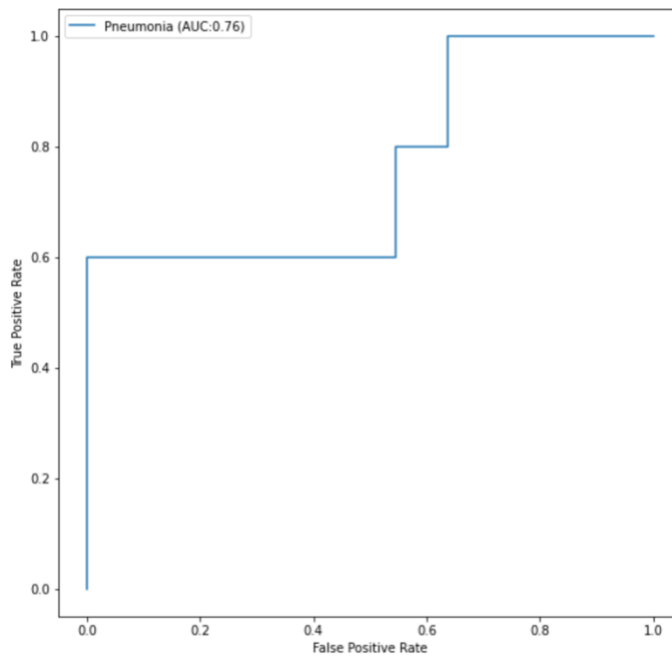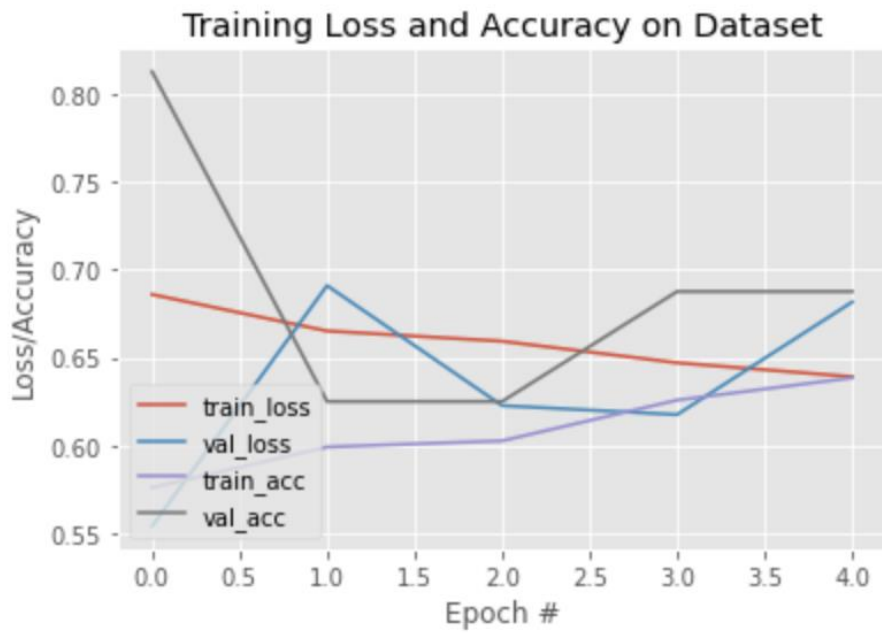
Final Threshold and Explanation:


Precision and Recall over Thresholds

Based on the above plot, the threshold is set as 0.25 which favors recall but also keeps a good enough precision.

4. Databases

Description of Training Dataset: There are 2290 rows and 27 columns in our training dataset. The pneumonia_class is balanced with 50 percent of the data of patients with pneumonia.
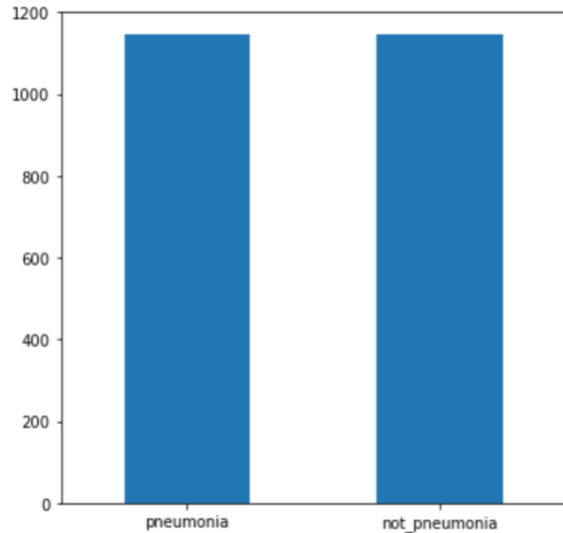
Description of Validation Dataset: There are 1430 rows and 27 columns in our testing. The pneumonia_class is imbalanced with 80 percent of the data of patients with pneumonia.
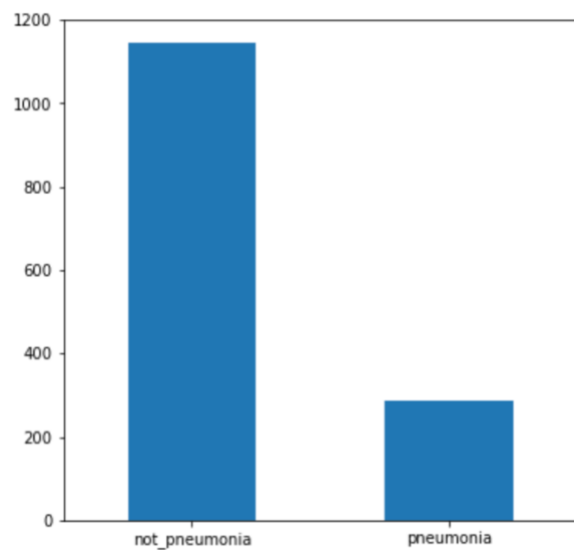
5. Ground Truth

The labels are obtained using an NLP approach from the radiologist reports.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

The dataset consists of women and men ages between 10 and 80 years with no history of previous Pneumonia disease.

Ground Truth Acquisition Methodology:

NLP-derived labels

Algorithm Performance Standard:

The algorithm performance standard is f1-score. The minimum acceptable f1-score is 0.387 which is stated in the paper in the following link. https://arxiv.org/pdf/1711.05225.pdf