FDA Submission

Your Name:
Hui Ren

Name of your Device:
Pneumonia detector

Algorithm Description
I used the CNN architecture and transfer learning to make binary prediction.

1. General Information

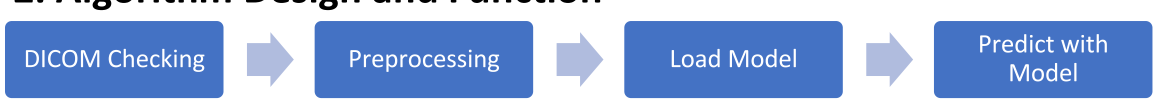
Intended Use Statement:
Assisting the radiological diagnosis of pneumonia from chest X-rays with the view positions of AP and PA.

Indications for Use:
Shortening the turn-around time of radiological diagnosis for males and females from 1 to 100 years of age with a chest X-ray. The male to female ratio is about 1.2. The comorbidities that the indicated population could exhibit include Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural_Thickening, and Pneumothorax.

Device Limitations:
Requiring high power processing such as GPU and cloud infrastructure.

Clinical Impact of Performance:
False negatives mean that the patients with pneumonia will not receive the treatment. False positives mean that the patients without pneumonia will receive the treatment. False negatives in this scenario are worse because patient's health is likely to deteriorate without treatment.

2. Algorithm Design and Function



DICOM Checking Steps:
I used the pydicom library to obtain the pixel_array, modality, patient position, body part examined and findings. I perform a check to see if the modality is "DX", the patient position is "AP" or "PA" and the body part examined is "CHEST". If the DICOM does not meet all the criteria, the image will not be assessed.

Preprocessing Steps:

I rescaled the pixel_array by dividing it with 255.0. I also stacked and resized the pixel_array to fit the input shape of the model (1, 224, 224, 3).

CNN Architecture:

I used the CNN architecture from VGG16.

3. Algorithm Training

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
global_average_pooling2d_1 (	(None, 512)	0
dense_1 (Dense)	(None, 1)	513

Parameters:
Types of augmentation used during training: horizontal flip, height shift, width shift, rotation, shear and zoom.

Batch size:
30

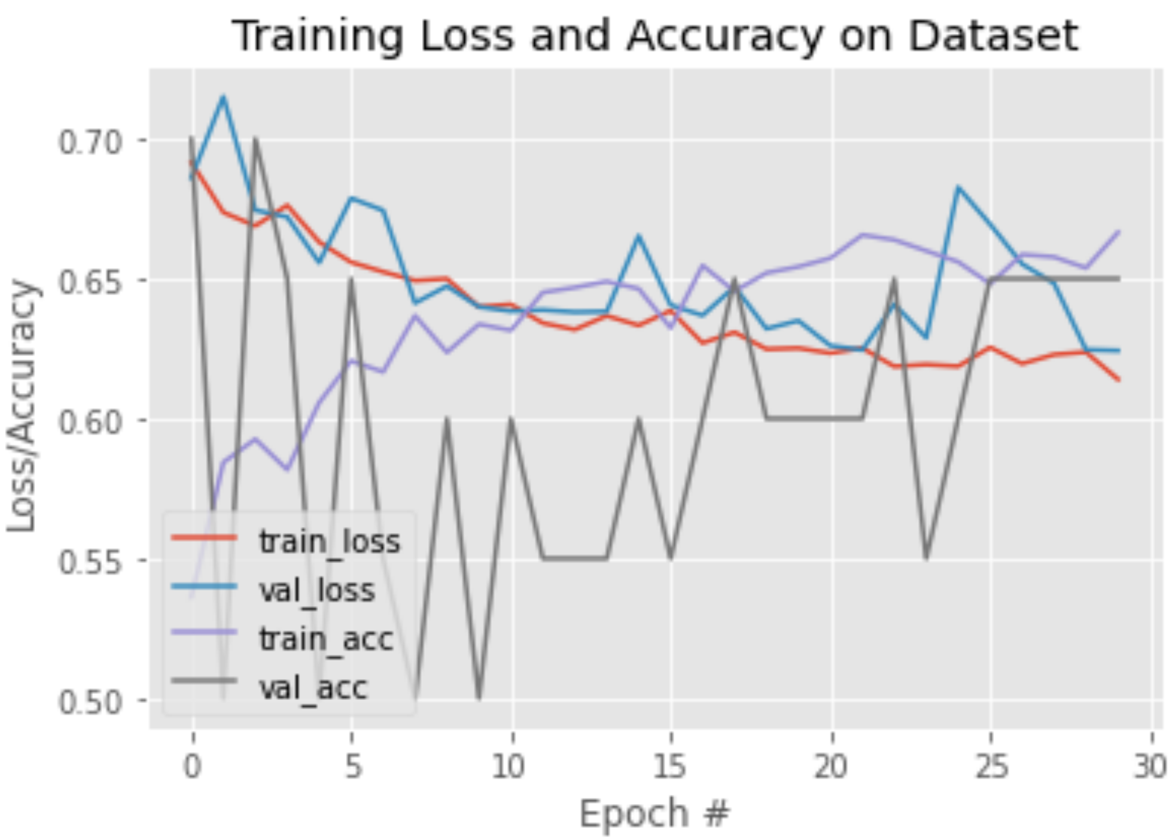
Optimizer learning rate:
1e-4 with a Learning Rate Scheduler that drops the learning rate in half every 10 epochs.

Layers of pre-existing architecture that were frozen:
The first 17 layers were frozen.

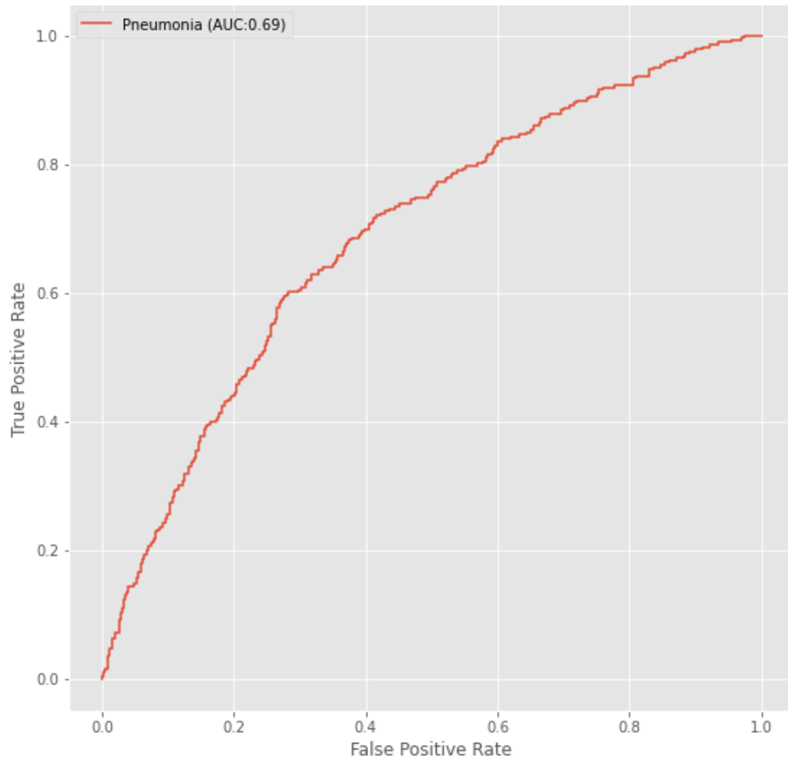
Layers of pre-existing architecture that were fine-tuned:
Block5_conv3 and block5_pool.

Layers added to pre-existing architecture:
Global average pooling layer and one dense layer.

Algorithm training performance visualization

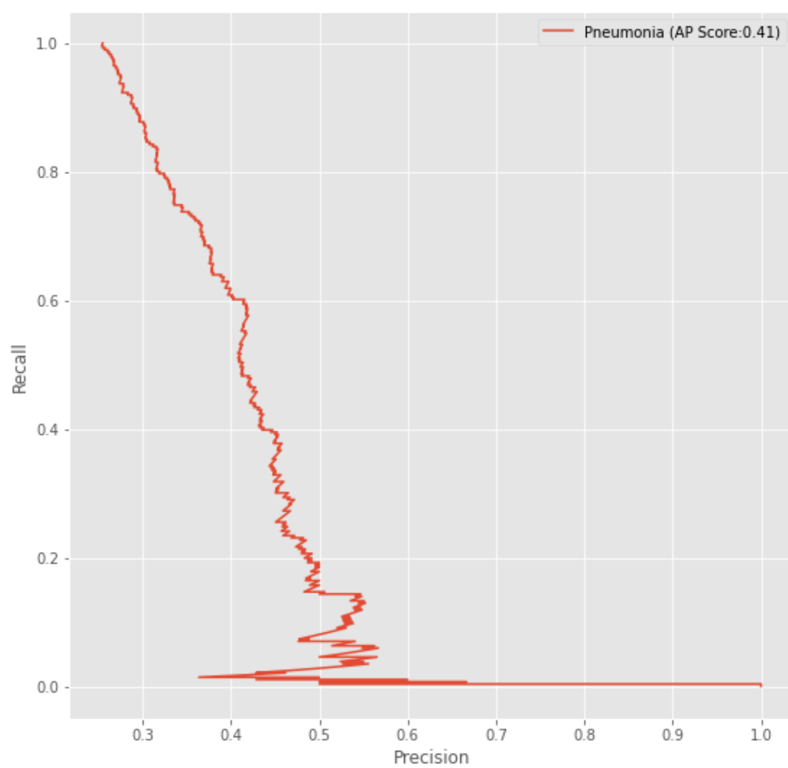


FP-RP curve

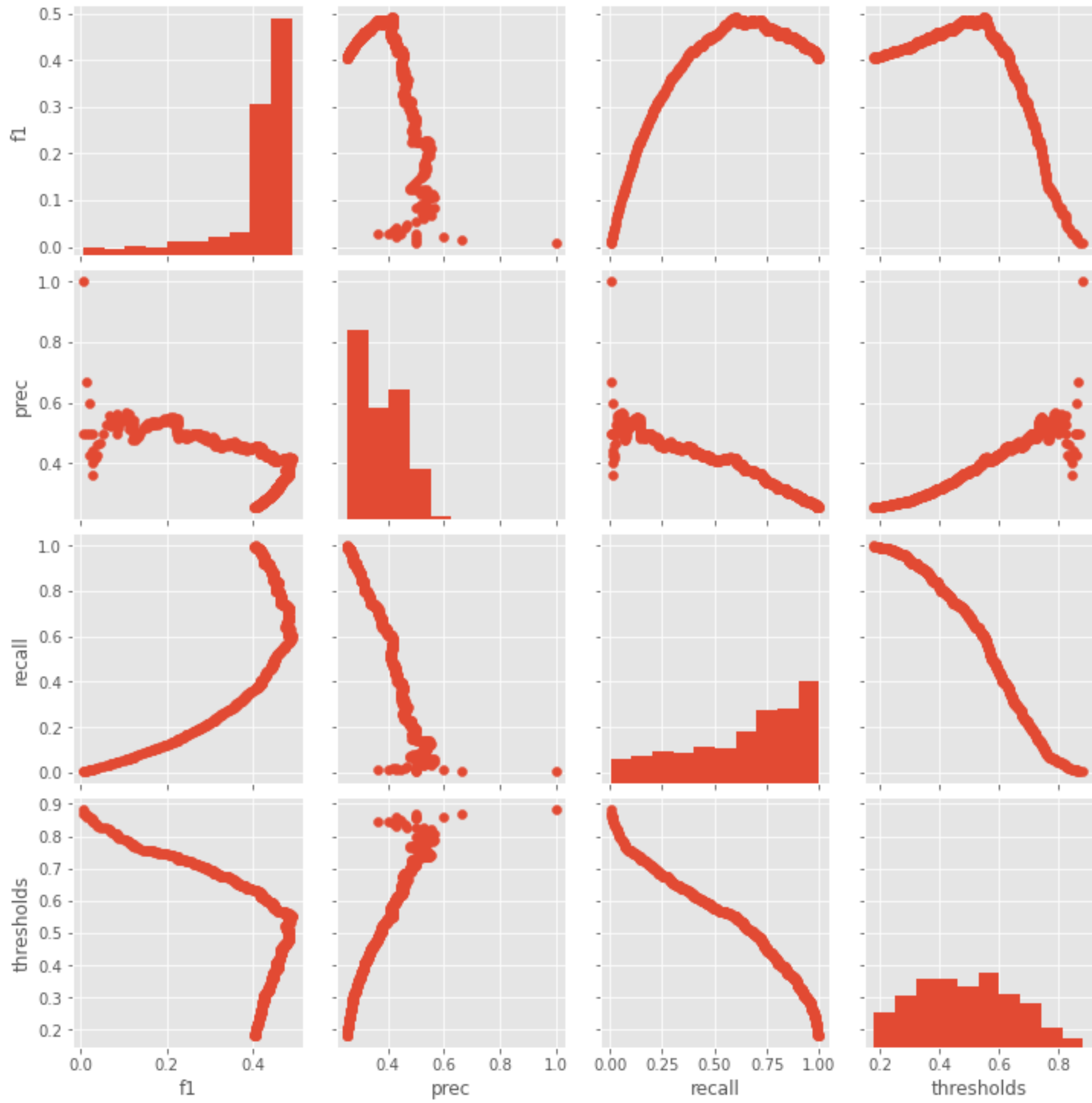


P-R curve

Therefore, I chose F1 as the performance standard to balance precision and recall in the clinical task of diagnosis.



Final Threshold and Explanation:
0.550 gives the maximum f-beta score. It balances both precision and recall.



4. Databases

Both the training and validation datasets are part of the NIH Chest X-ray Dataset.

Description of Training Dataset:
I performed class balance to generate the equal number of positive and negative cases.

Description of Validation Dataset:

I performed class balance to generate a 25% of positive cases, to align with the percentage in the clinical settings.

5. Ground Truth

The ground truth is NLP-derived labels. NLP at this stage is not sophisticated enough to understand the nuances of the natural languages. The rough estimate is that the accuracy would be around 90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

Males and females with age from 1 to 100. The male to female ratio is about 1.2. The comorbidities include Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural_Thickening, and Pneumothorax. In the DICOM header of the X-ray, the modality should be "DX", the patient position should be "AP" or "PA" and the body part examined should be "CHEST".

Ground Truth Acquisition Methodology:

The silver standard of radiologist reading.

Algorithm Performance Standard:

The F1 score is about 0.435 as indicated by [this] paper.

	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)