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, image type, patient position body part and findings.

Preprocessing Steps:
I normalized the pixel_array by the mean and standard deviation. I also resized the pixel_array.

CNN Architecture:
I used the CNN architecture from VGG16.

3. Algorithm Training

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
model_1 (Model)              (None, 7, 7, 512)         14714688
_________________________________________________________________
global_average_pooling2d_1 ( (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 513
=================================================================
Total params: 14,715,201
Trainable params: 2,360,321
Non-trainable params: 12,354,880
```

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

Batch size:
15

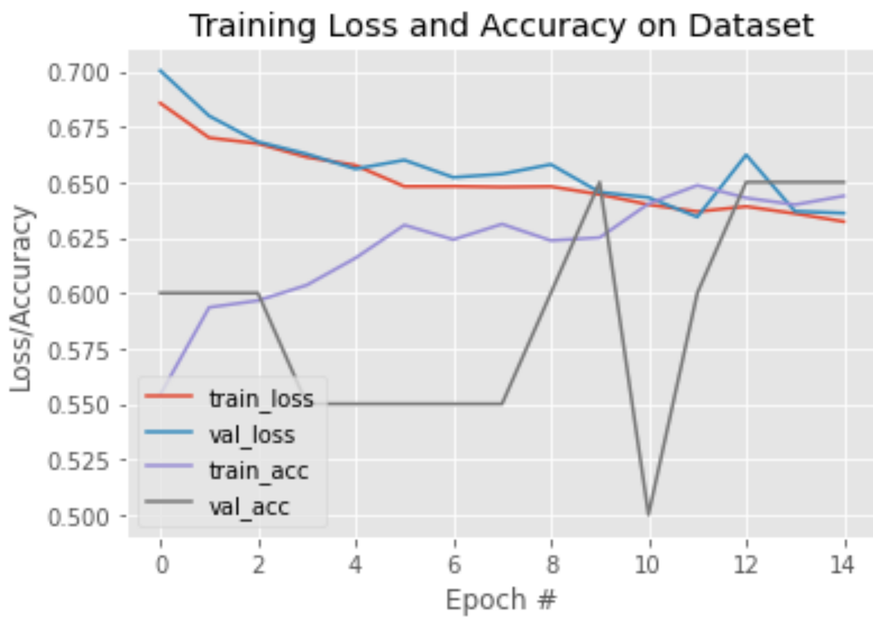
Optimizer learning rate:
1e-4

Layers of pre-existing architecture that were frozen:
The first 16 layers were frozon.

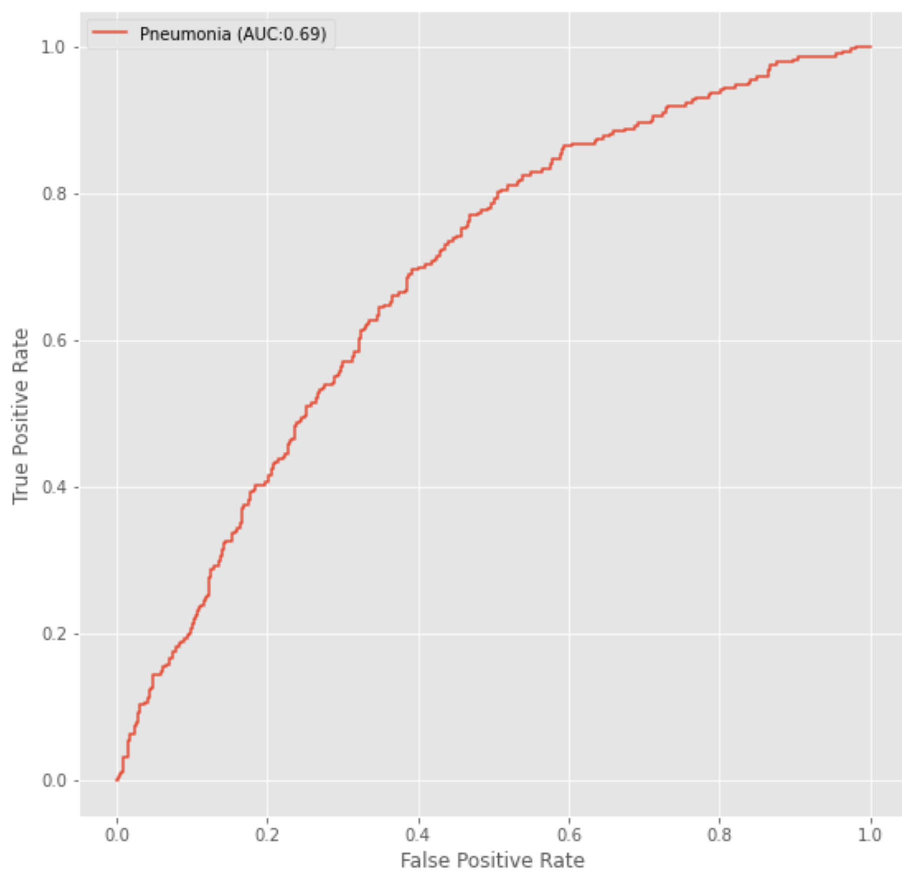
Layers of pre-existing architecture that were fine-tuned:
Block5_conv3 and block5_pool.

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

Algorithm training performance visualization

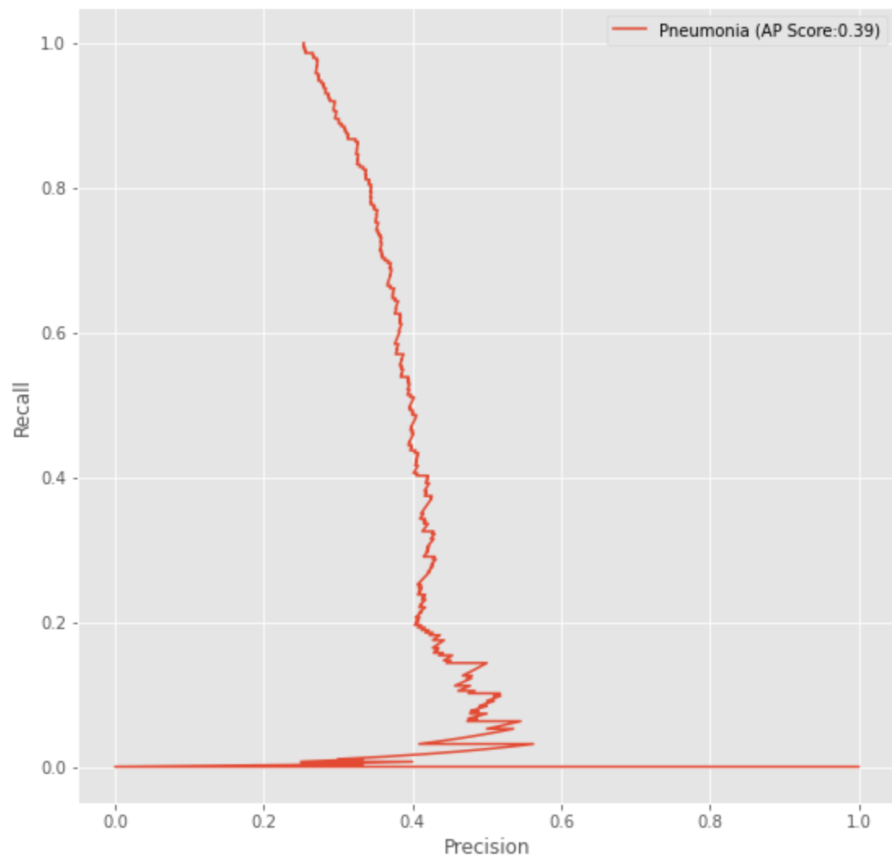


FP-RP curve

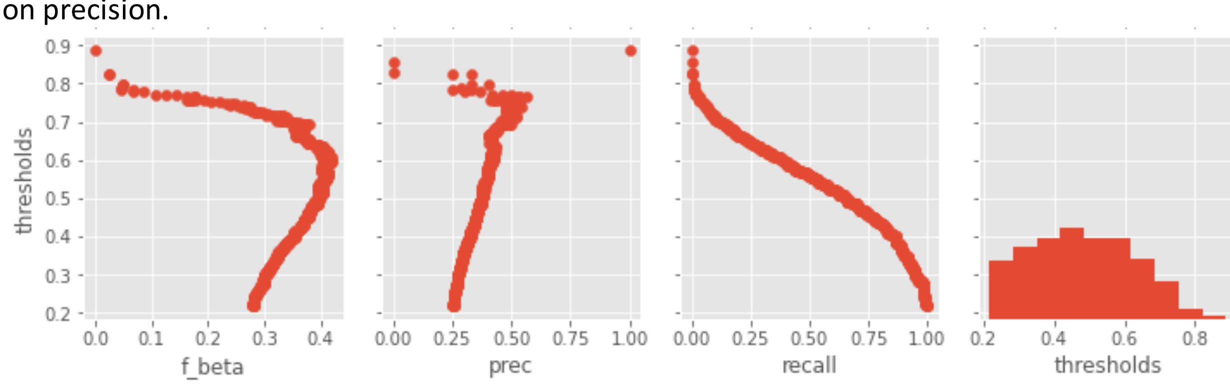


P-R curve

Precision in this scenario is more important than recall due to the clinical task of diagnosis. Therefore, I chose F-beta as the performance standard with the beta score smaller than 1.



Final Threshold and Explanation:
0.827 gives the maximum f-beta score. It balances both precision and recall, with the emphasis on precision.



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.

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)