

FDA Submission

Your Name:

Donald Doran

Name of your Device:

pneuDetector

Algorithm Description

1. General Information

Intended Use Statement:

pneuDetector is intended used for assist the radiologists to detect pneumonia from chest X-ray images for the patient of ages 1 to 90.

Indications for Use:

pneuDetector can analyse chest X-ray image in 'AP' or 'PA' position from people of ages 1 to 90 by CNN. It can be used to help the radiologists to detect pneumonia.

pneuDetector can be used to optimize the workflow of pneumonia detection by radiologists. It generate predicated label and a score (0 ~ 1) to indicate the likelihood of pneumonia detection. It could shorten patient's waiting time, and reduce radiologists' workload.

Device Limitations:

pneuDetector is used for pneumonia screening rather than giving the final diagnosis. The result of pneuDetector can not be used for diagnosis alone. The result of pneuDetector should be used in combination with reports from radiologists, the laboratory test results, the health record of the patient and other information to conclude the final diagnostic result.

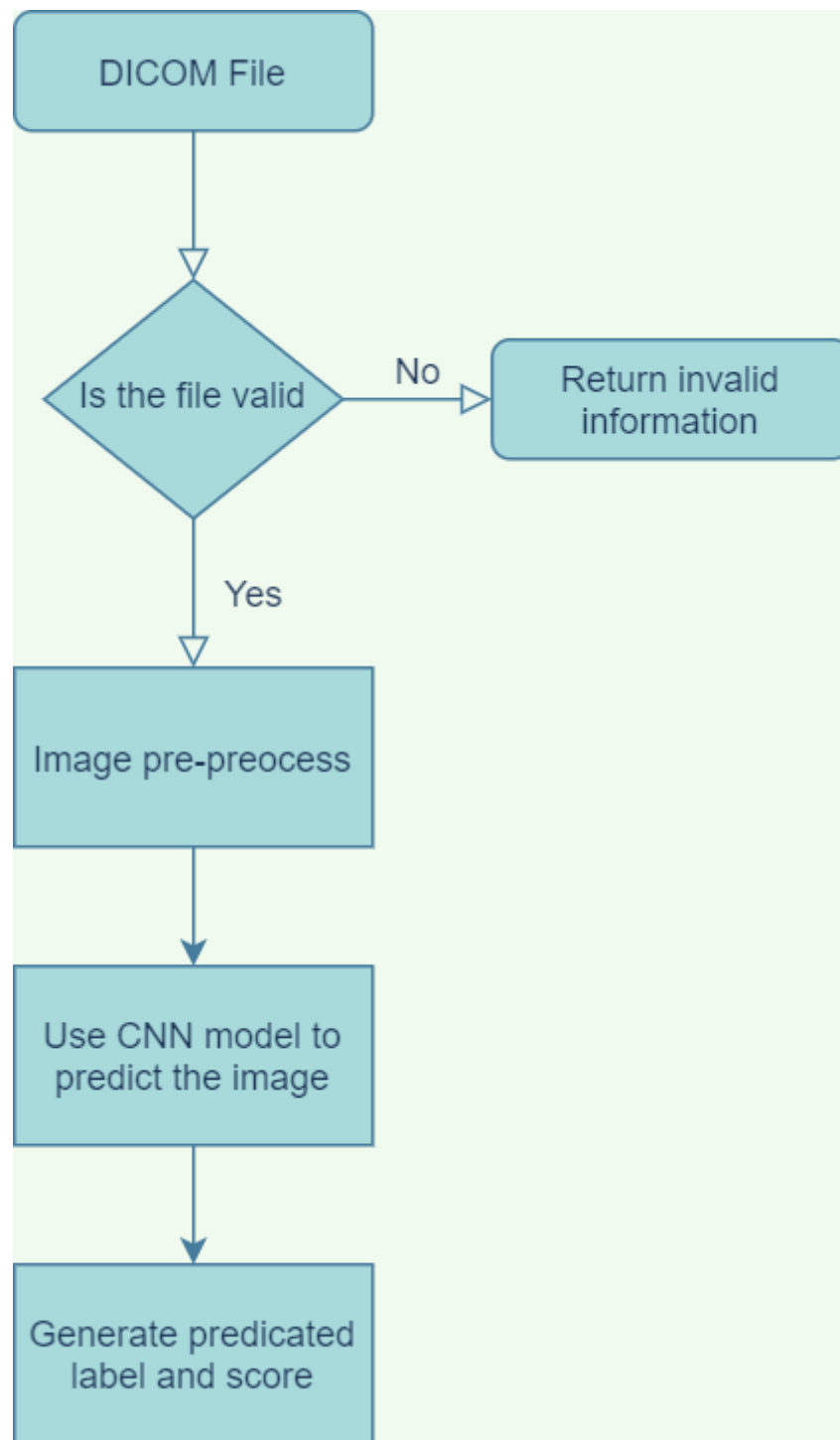
Clinical Impact of Performance:

A False Negative result will mislabel the 'Pneumonia' case as 'No Finding', and may delay the diagnosis of the patient.

A False Positive will mislabel the image of normal or other disease as 'Pneumonia'. Since our algorithm is only used to assist radiologists to screen pneumonia rather than give the final diagnosis, a False Positive result may increase the waiting time of the patient, but will not lead to the mis-diagnosis of the patient.

Therefore, a threshold that favors high False Positive number and low False Negative number was chosen.

2. Algorithm Design and Function



DICOM Checking Steps:

In this step the algorithm will check the flowing information in the DICOM file.

The input DICOM that does not meet the following criteria will be considered as valid input, and can not be analysed by pneuDetector.

- Modality should be `DX`
- Body Part Examined should be `CHEST`
- Patient Age should in the range of 1 ~ 90
- Patient Position should be `AP` or `PA`

Preprocessing Steps:

The image of the DICOM file will be used as the input of pneuDetector.

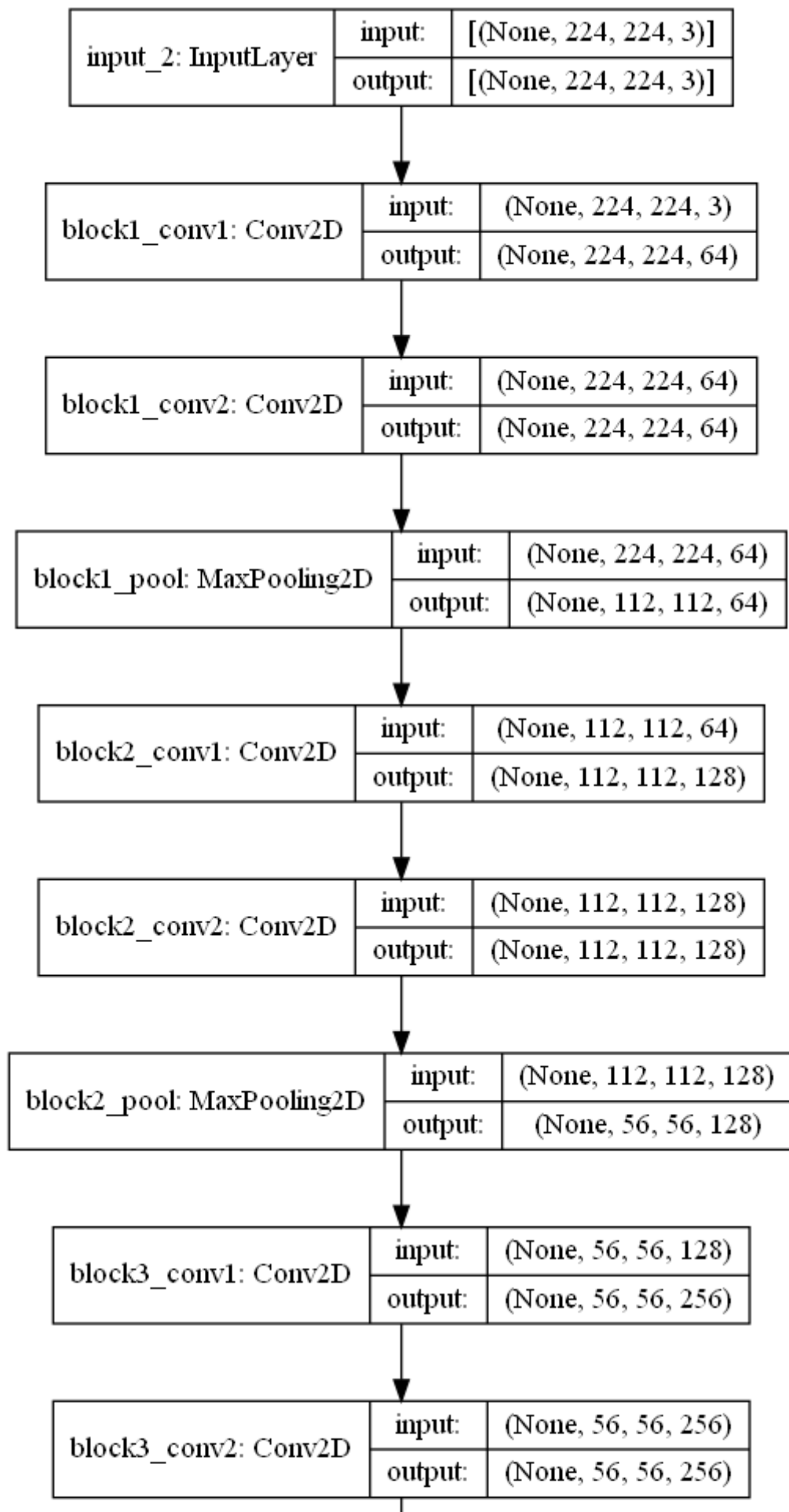
The image will be resized to (1 * 224 * 224 * 3), which is required by the following CNN Architecture.

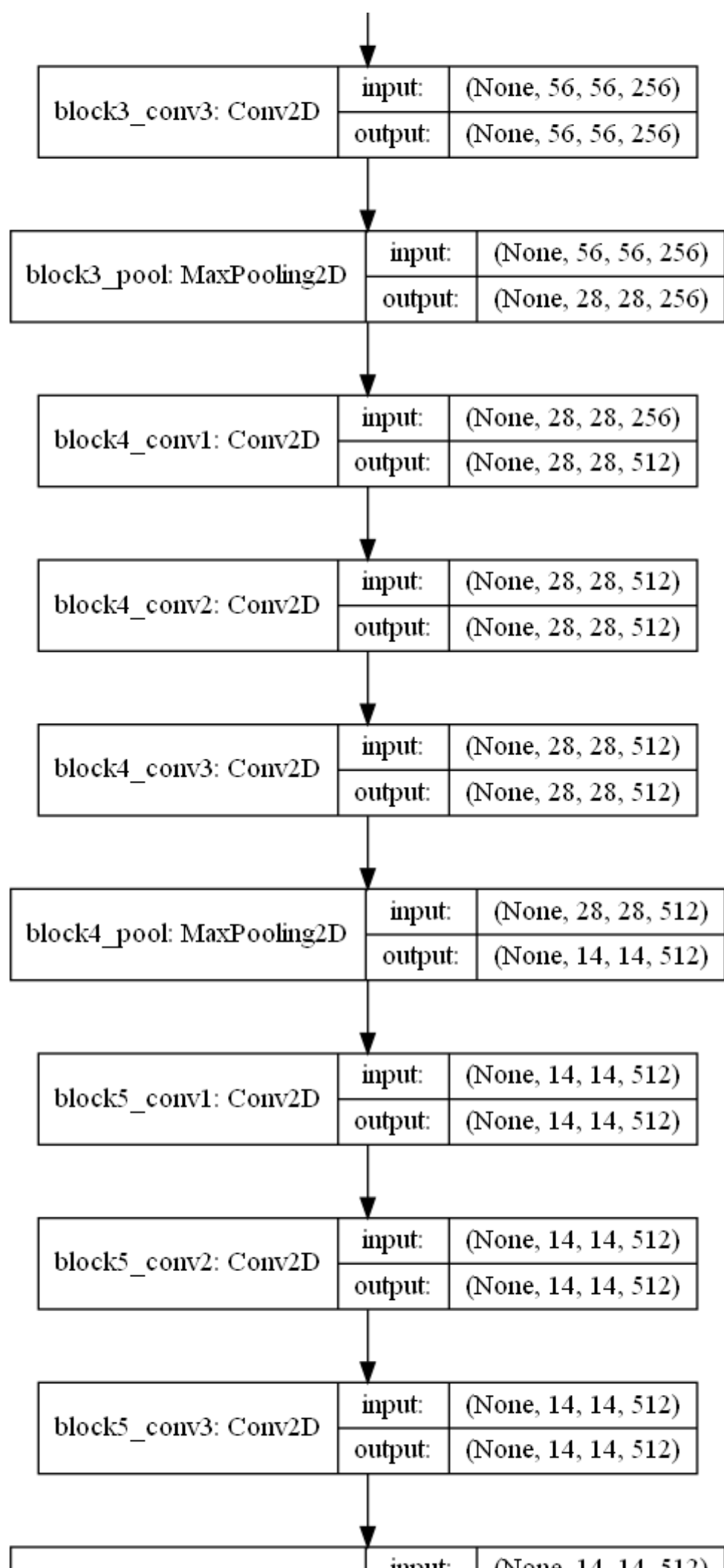
Then the resized image will rescale to 1/255 and normalized

CNN Architecture:

pneuDetector used the VGG16 in conjunction with additional fully connected layers to build a model.

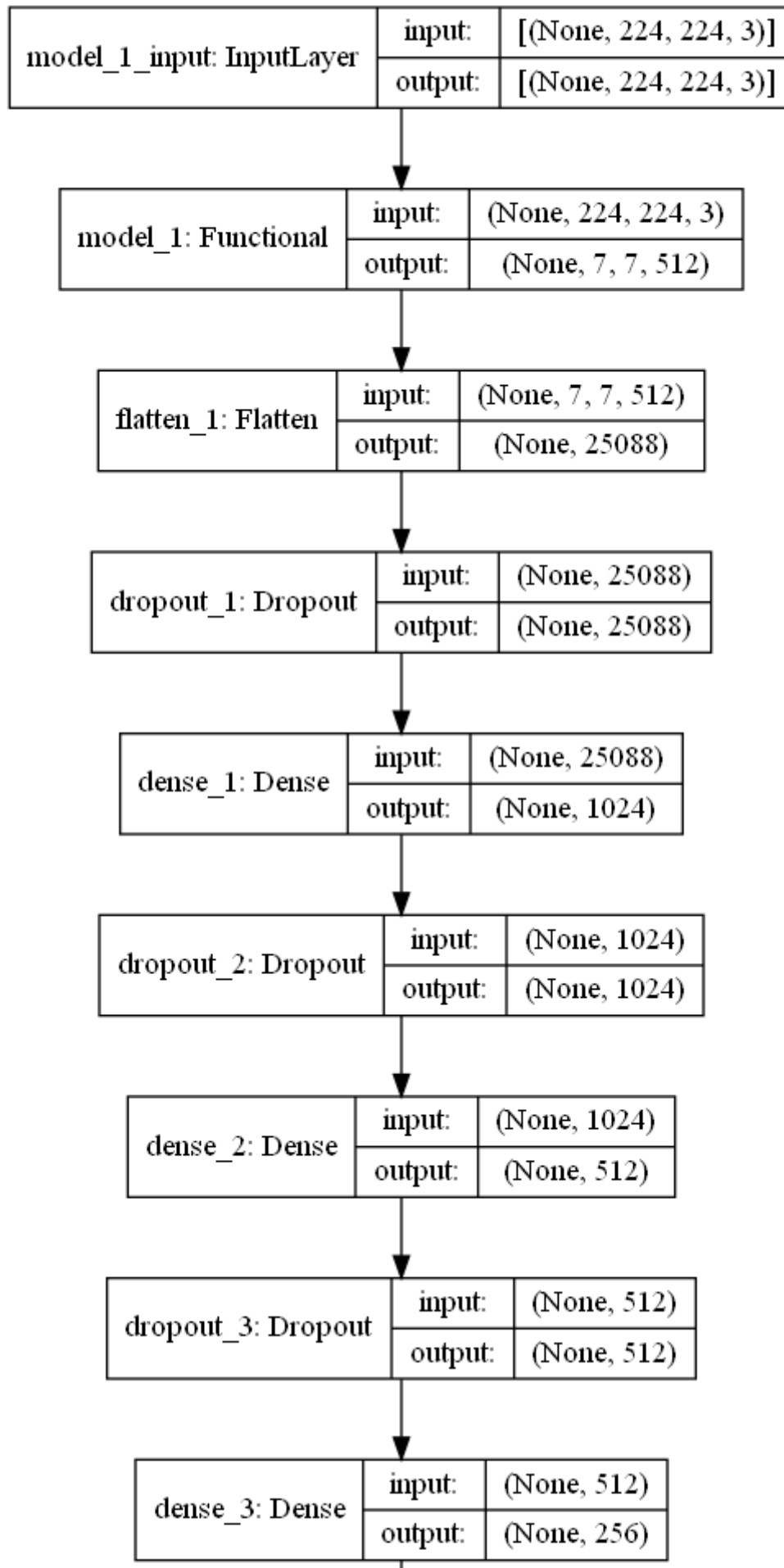
The Architecture of VGG16:

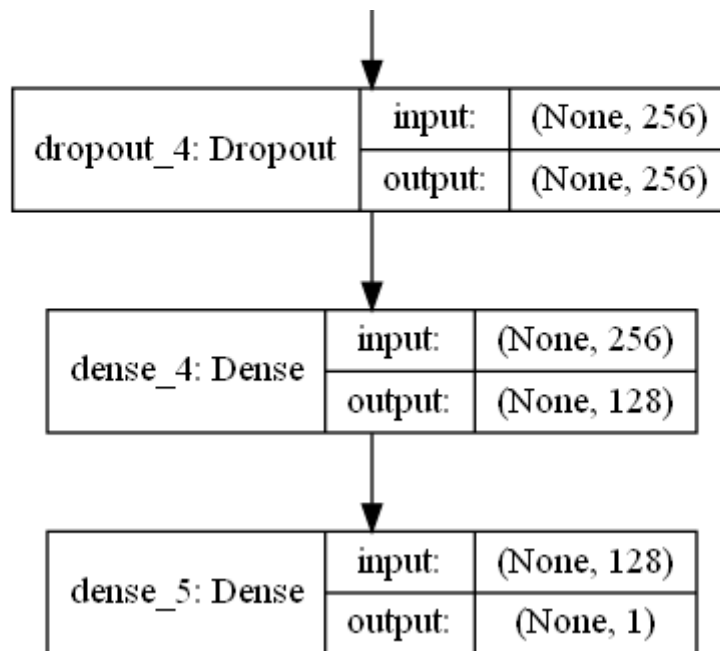




block5_pool: MaxPooling2D	input:	(None, 14, 14, 512)
	output:	(None, 7, 7, 512)

The Architecture of CNN:





3. Algorithm Training

Training DataSet Types of augmentation:

- rescale : 1. / 255.0
- samplewise_center : True
- samplewise_std_normalization : True
- horizontal_flip : True
- height_shift_range : 0.1
- width_shift_range : 0.1
- rotation_range : 10
- shear_range : 0.1
- zoom_range : 0.1

Validation and Test DataSet types of augmentation:

- rescale : 1/ 255.0,
- samplewise_center : True,
- samplewise_std_normalization : True

Batch size:

- training batch size: 24
- validation batch size : 50
- validation_steps : 12

Training Process Parameters:

- Optimizer learning rate : Adam(lr : 1e-4)
- Loss Function: binary_crossentropy
- Metrics: accuracy
- Layers of pre-existing architecture that were frozen : 0 - 16
- Layers of pre-existing architecture that were fine-tuned : 17 - 18

Layers added to pre-existing architecture

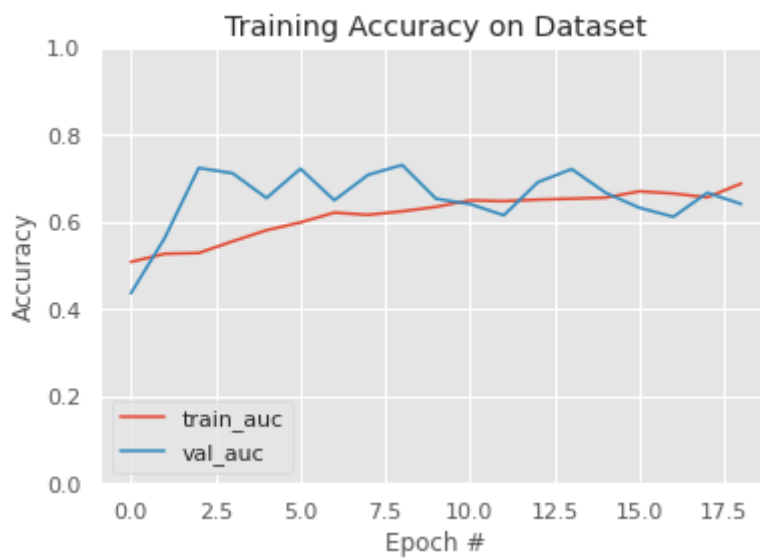
- flatten_1 (Flatten)
- dropout_1 (Dropout)
- dense_1 (Dense)

- dropout_2 (Dropout)
- dense_2 (Dense)
- dropout_3 (Dropout)
- dense_3 (Dense)
- dropout_4 (Dropout)
- dense_4 (Dense)
- dense_5 (Dense)

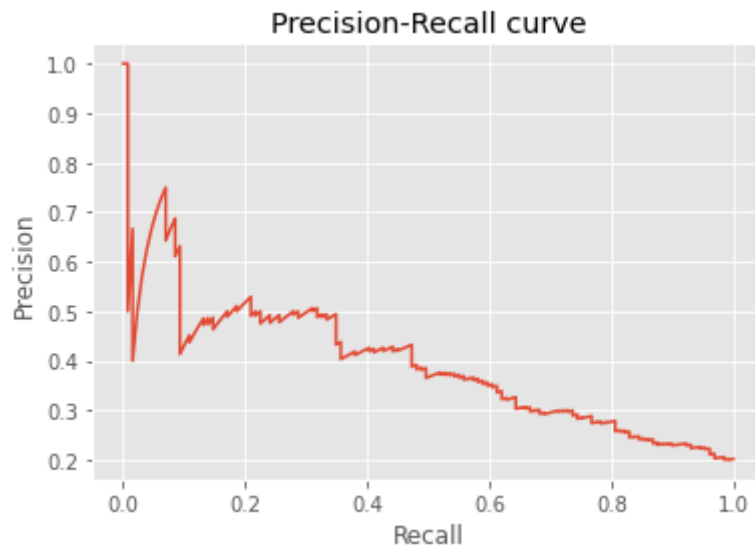
Training Loss of loss



Training Loss of accuracy

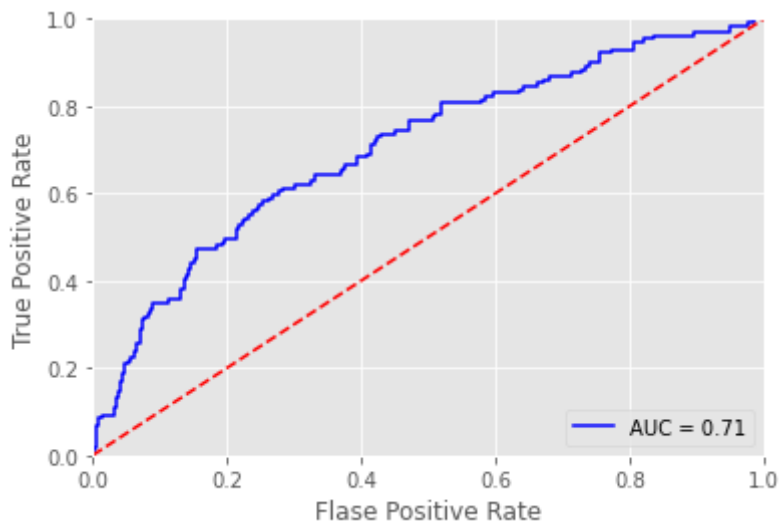


Precision Recall Curve



Final Threshold and Explanation:

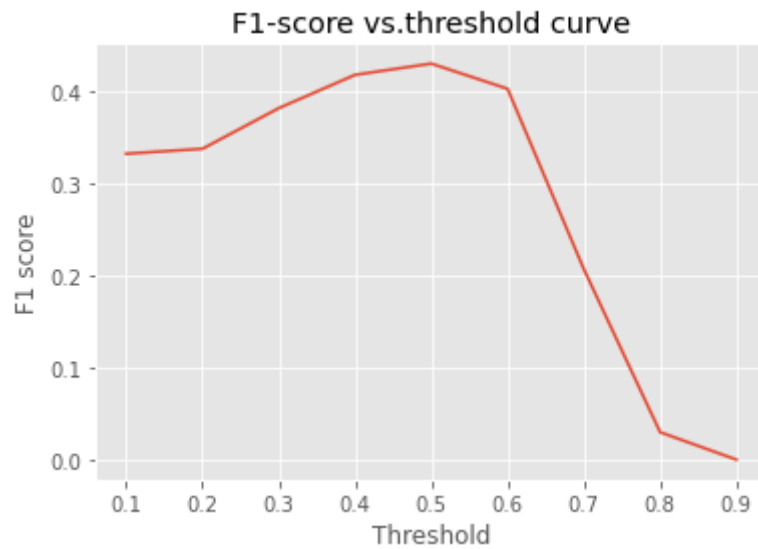
ROC Curve



Model Performance Evaluation

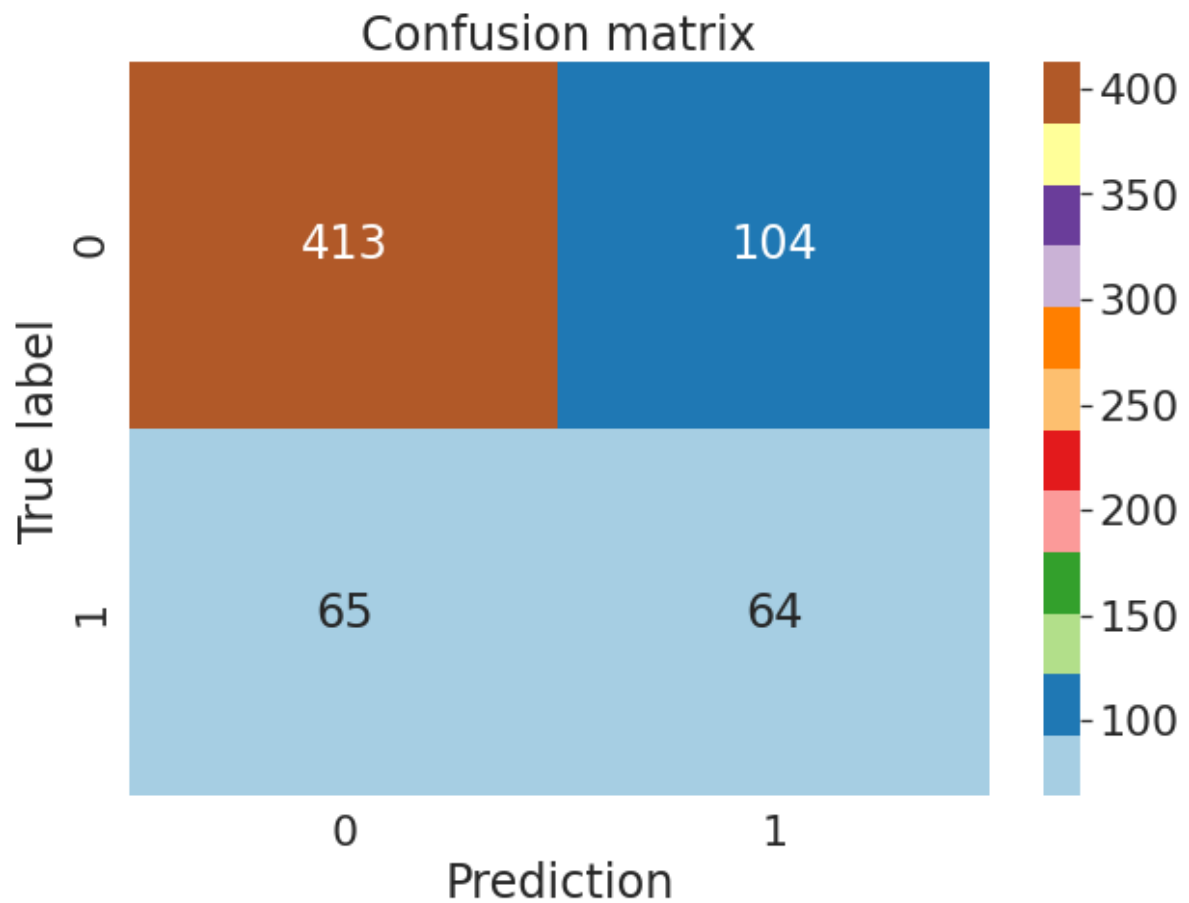
Threshold	Accuracy	Precision	Recall	F1
0.1	0.199690	0.199690	1.000000	0.332903
0.2	0.243034	0.204918	0.968992	0.338295
0.3	0.455108	0.247166	0.844961	0.382456
0.4	0.634675	0.306859	0.658915	0.418719
0.5	0.738390	0.380952	0.496124	0.430976
0.6	0.794118	0.478723	0.348837	0.403587
0.7	0.798762	0.485714	0.131783	0.207317
0.8	0.798762	0.400000	0.015504	0.029851
0.9	0.800310	1.000000	0.000000	0.000000

F1-score vs.threshold curve:



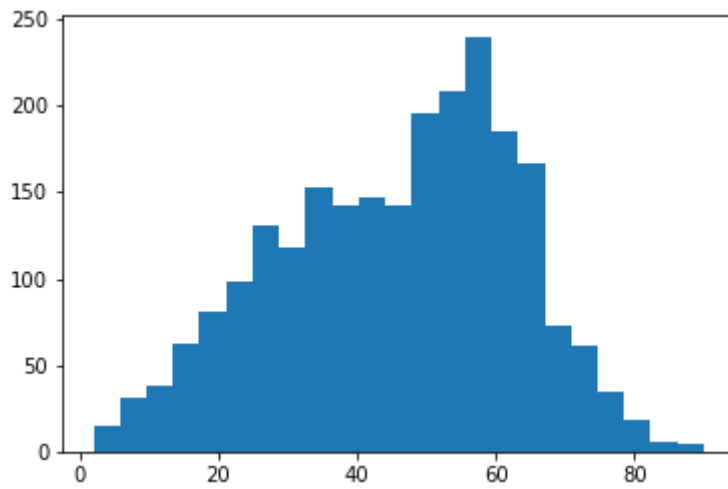
pneuDetector is used to assist the diagnosis of pneumonia. Because F1 score combines both precision and recall and allows us to better measure a test's accuracy when there are class imbalances. We choose 0.5 as threshold with the highest F1 score(0.43).

Confusion Matrix(threshold : 0.5):

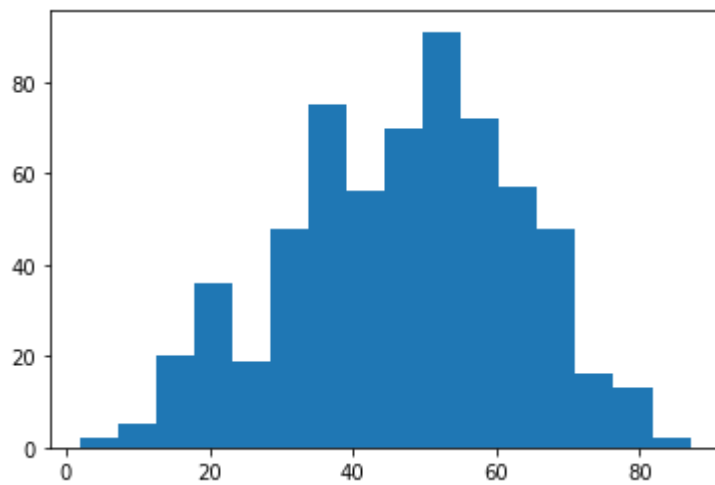


4. Databases

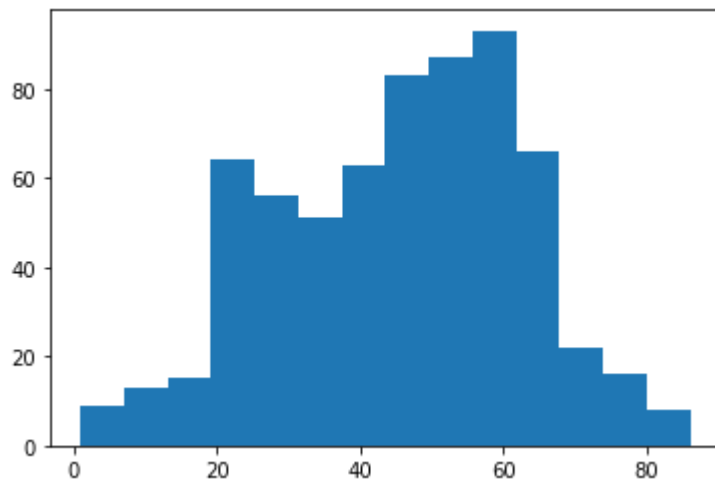
Age Distribution of Training Dataset:



Age Distribution of Validation Dataset:



Age Distribution of Test Dataset:



Dataset	Training	Validation	Test
Size	2350	630	646
Pneumonia Positive Rate	0.5	0.2	0.2
Male/Female	0.58	0.55	0.57
Age(Max)	90	87	86
Age(Min)	2	2	1
Age(Mean)	46	47	46

5. Ground Truth

The dataset used for training the algorithm is from NIH Chest X-ray Dataset.

ChestX-ray dataset comprises 112,120 frontal-view X-ray images of 30,805 unique patients with the text-mined fourteen disease image labels, mined from the associated radiological reports using natural language processing.

Because the original radiology reports are not publicly available. the NLP mined labels were used as ground truth for the dataset. The NLP labelling accuracy is estimated to be >90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

Ages of 1 to 90, with suspicious for pneumonia. Chest X-ray image should be taken at 'AP' or 'PA'.

Ground Truth Acquisition Methodology:

The golden standard to acquire the ground truth is by laboratory testing or from expert labeling, they are both expensive and hard to obtain.

Because our algorithm is used for assist the radiologists, we will obtain the ground truth by the silver stanard.

We will hiring several radiologists to lable the images, and make the final diagnosis by a voting system from their labels.

Algorithm Performance Standard:

The algorithm performance can be measured by the F1 score. As some article have suggested, the average F1 score of radiologists is 0.387. ([CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rayswith Deep Learning](#)).

In order to assist radiologist, our algorithm should get a F1 score equal or higher than 0.387.