FDA Submission

Your Name: Slaouti-Jégou Yannis

Name of your Device: Pneumonia Detection Algorithm for Chest X-Rays

Algorithm Description

1. General Information

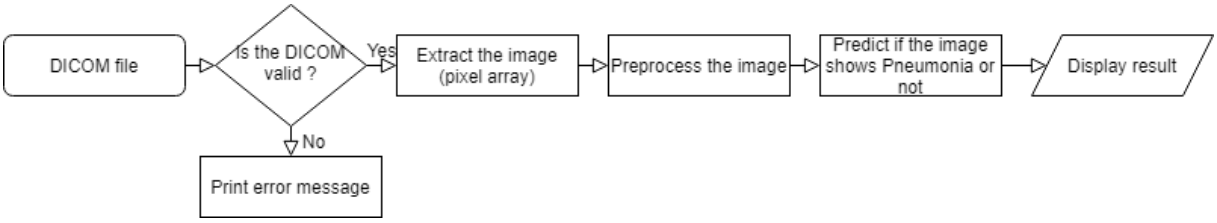
Intended Use Statement: For assisting the radiologist in the detection of pneumonia in chest x_rays.

Indications for Use: This algorithm is intended for use on men and women from the ages of 20-70 who have been administered a chest x-ray screening. This algorithm has a specificity around 0.8 so it should be used for ruling in pneumonia.

Device Limitations: The algorithm is not trained to perform with the presence of effusion in the x-ray. Also, most patients used in the training were between the age of 20 to 70.

Clinical Impact of Performance: The algorithm does not require to be run in an emergency setting so it does not require high-performance computing. False positives might lead to an unnecessary doctor check-up and false negatives are dangerous which is why at least one radiologist should to go through all the x-rays even when labeled negative by the algorithm.

2. Algorithm Design and Function



DICOM Checking Steps: Check that the image is a digital radiography (DX), check that it is a chest-focused x-ray and check that the patient position is either AP or PA.

Preprocessing Steps: Resize the image in the (1, 224, 224, 3) shape, standardize, zero-mean and rescale it (1/255).

CNN Architecture:

The architecture used is a fined-tuned VGG16 model. The transfer layer is the last 2D maxpooling layer and the layers added after that are the flatten layer followed by a combination of dense and dropout layers. The dropout layers prevent overfitting (dropout rate=0.5). The first 17 layers of the model are not trainable.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
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
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              25691136
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 512)               524800
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_3 (Dense)              (None, 256)               131328
_________________________________________________________________
dropout_3 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 257
=================================================================
Total params: 41,062,209
Trainable params: 28,707,329
Non-trainable params: 12,354,880
```

3. Algorithm Training

Parameters:

- Types of augmentation used during training:
  ImageDataGenerator augmentations
  - Rescale (1. / 255.0), standardize and zero-mean
  - horizontal_flip
  - height shift range (0.1)
  - width shift range (0.1)
  - rotation range (15°)
  - shear range (0.1)
  - zoom range (0.1)

  custom contrast stretch from skimage rescale_instensity with percentiles 3, 97.

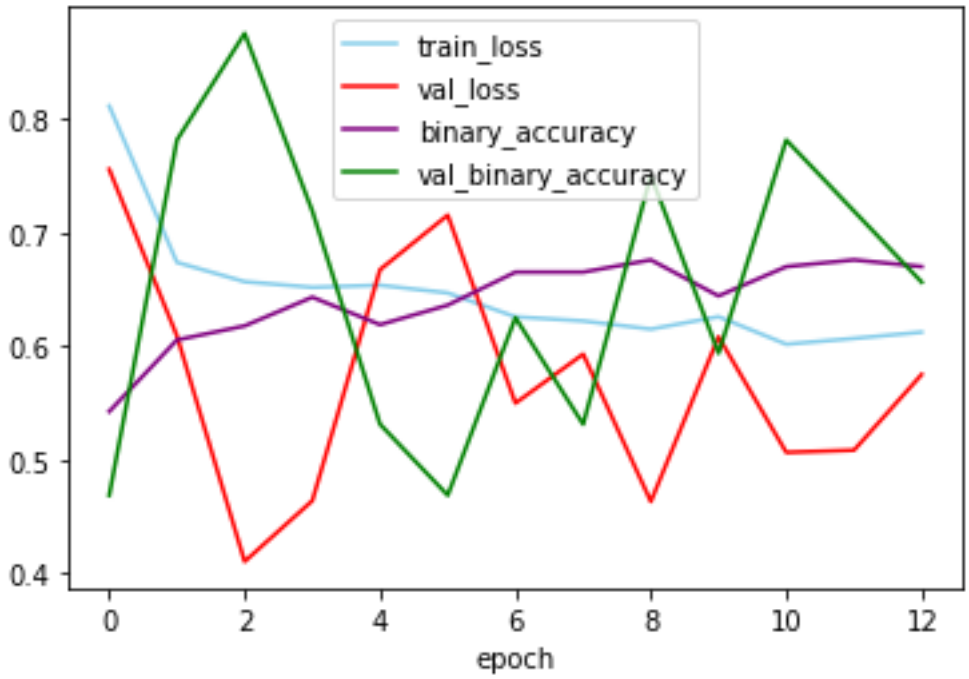
- Batch size: 32

- Optimizer learning rate: 1e-3

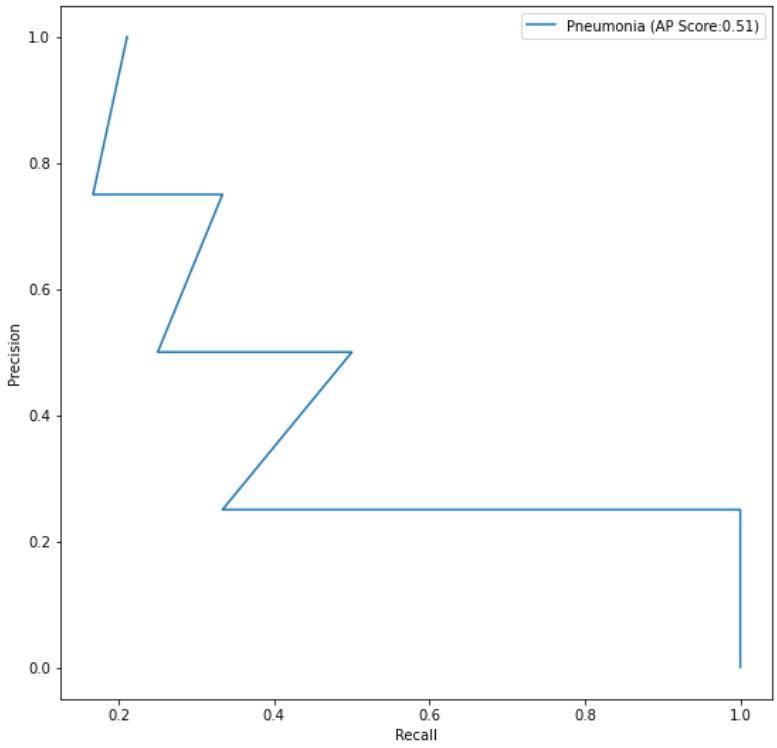
- Layers of pre-existing architecture that were frozen: 17

- Layers of pre-existing architecture that were fine-tuned: 2

- Layers added to pre-existing architecture: 8



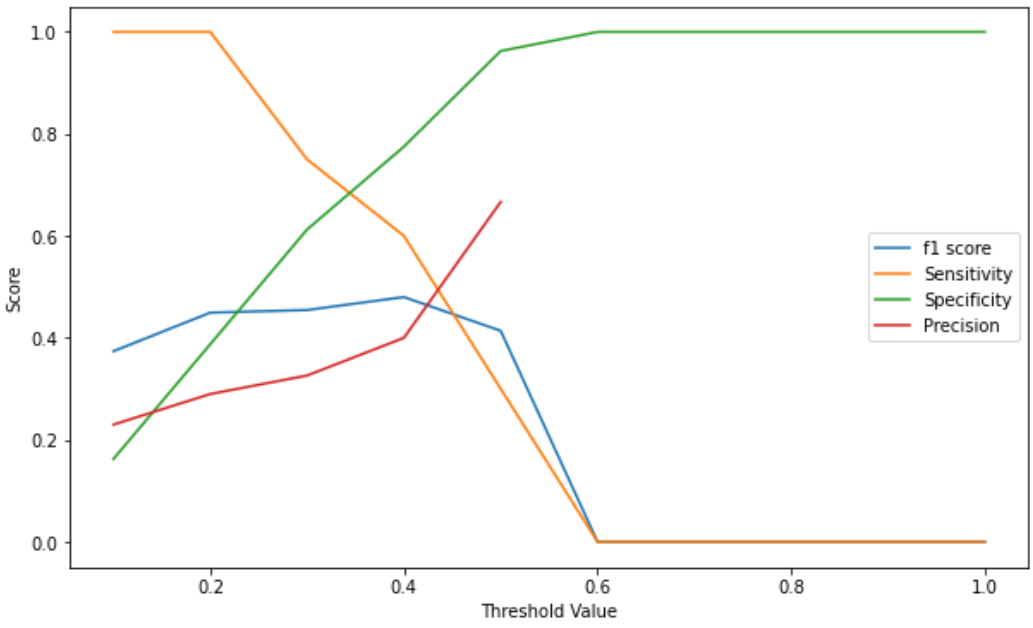
Training Performance



Precision/Recall

Final Threshold and Explanation: 0.4

- The chosen threshold is 0.4 because it is the optimal threshold for the f1 score and has both a good recall (sensitivity) and specificity. A lower threshold would be better for recall but as false positives would increase it would be worst for both precision and specificity. It means that if we want to rule out pneumonia, it would better to use a lower threshold and get a better recall. But, if we want to rule in pneumonia, we have to increase the threshold (0.4+) and thus the false negatives. That way we decrease false positives and get both a good precision and a good specificity.

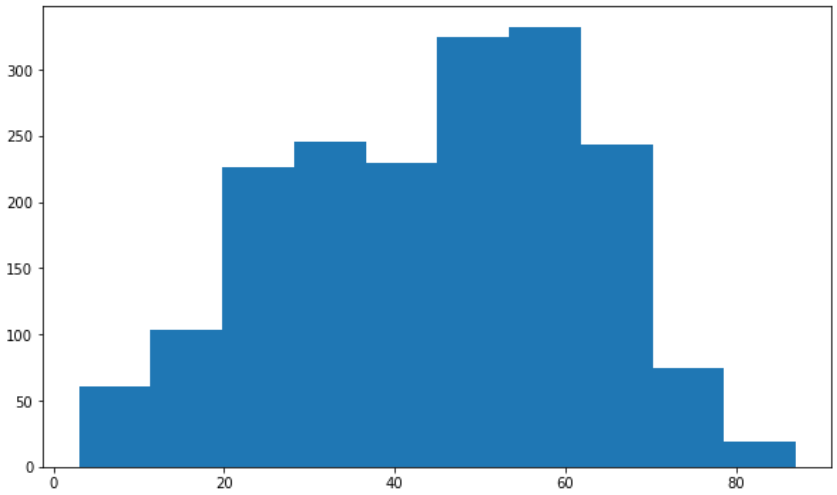


4. Databases

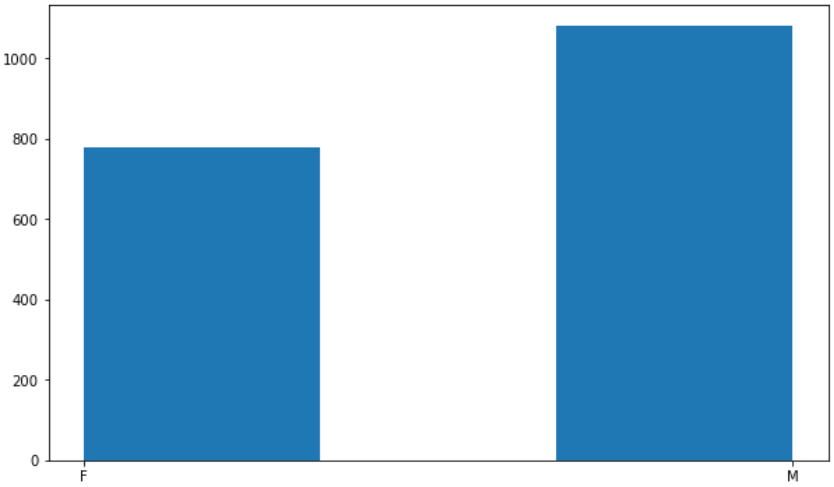
Description of Training Dataset:

The training set has 50% positive pneumonia cases out of a total of 1960 cases.

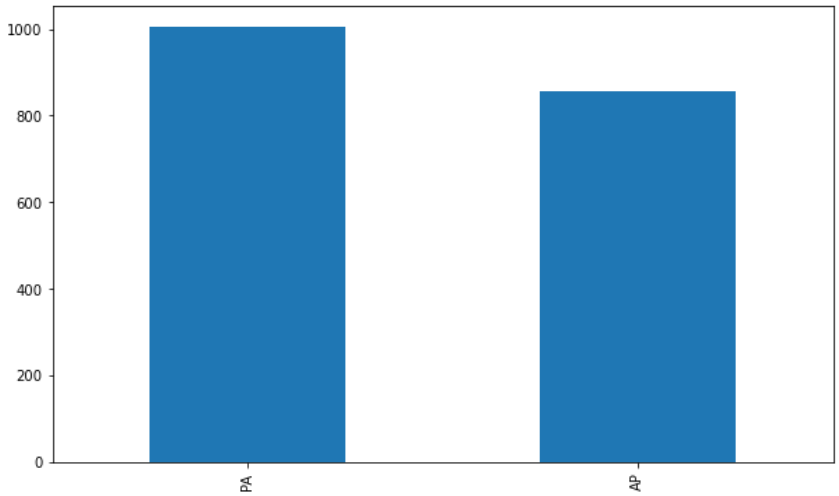
- Age-wise, most patients are between 20 and 70.



- Gender-wise, there are 1081 men and 779 women.



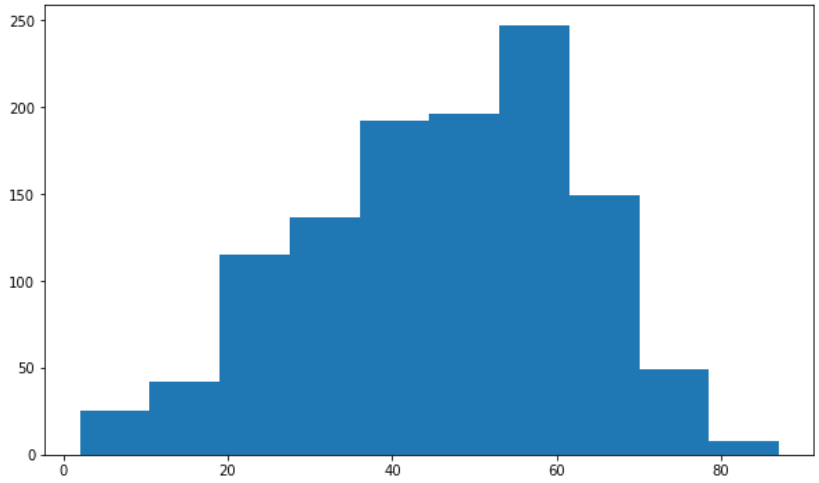
- For the patient position distribution, there are slightly more PA positions than AP.



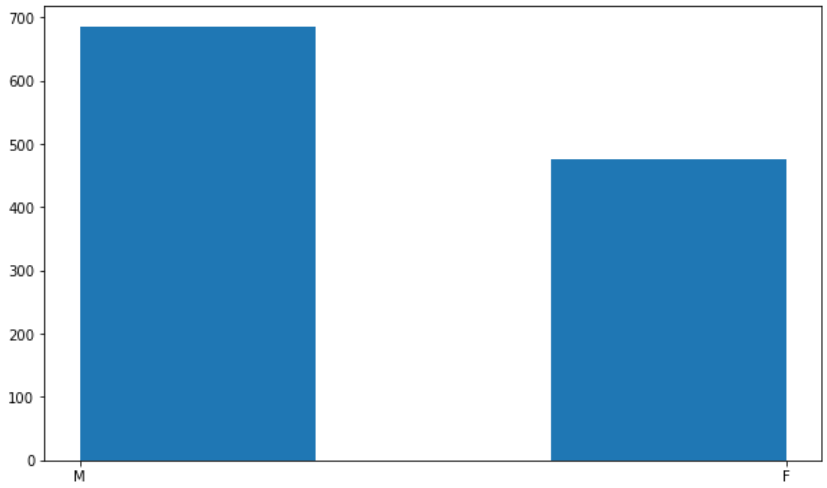
Description of Validation Dataset:

The validation set has 20% positive pneumonia cases out of a total of 1160 cases.

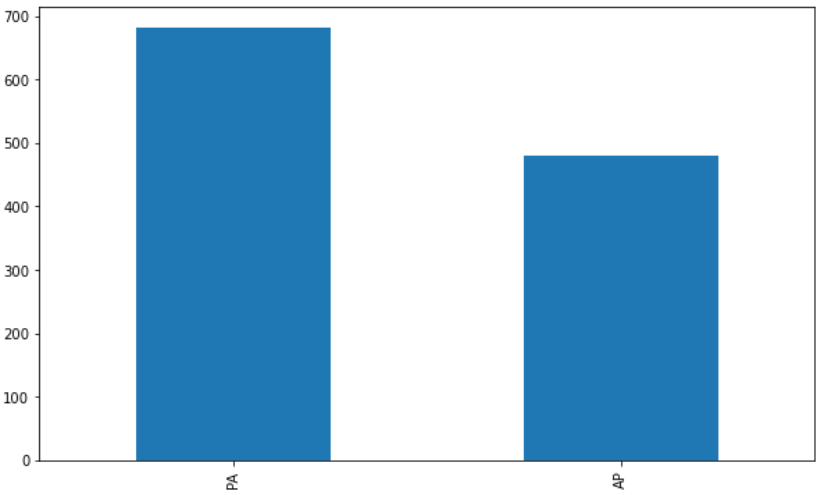
- Age-wise, most patients are between 20 and 70.



- Gender-wise, there are 685 men and 475 women.



- For the patient position distribution, there are more PA positions



5. Ground Truth

The ground truth labels were created using Natural Language Processing (NLP) to mine the associated radiological reports. The benefit of that is that we can obtain a large number of labels quickly, but the downside is that the accuracy of NLP is not perfect, and even though pneumonia is hard to detect we only have the expertise of a single radiologist.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

The ideal dataset I would want would be made up of chest x-ray screenings with a percentage of pneumonia that corresponds to the prevalence observed in the real world when there is a suspicion of pneumonia, so around 20-25%. Patients would be aged from 20 to 70, with a fairly equal gender distribution and a 60-40 AP-PA view positions distribution.

Ground Truth Acquisition Methodology:

The ideal ground truth acquisition would be for multiple radiologists (at least 3) to manually label the data. I would then apply a weighted mean of their labels based on their experience to determine how to apply the final ground truth label to an image.

Algorithm Performance Standard:

According to CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, the average f1-score for a radiologist when detecting pneumonia is 0.387. This is why I chose measure my algorithm's performance with the f1-score and it should have at least 0.4 as a f1-score standard.