FDA Submission

Your Name: Vishal Sharma

Name of your Device: X-Ray Pneumonia Detection Assist

Algorithm Description

1. General Information

Intended Use Statement:

This Algorithm is used to detect abnormalities in chest radio images provided, in order to help the radiologist make a decision on the presence or absence of pneumonia in the X-Ray of the patient (data from the EDA).

Indications for Use:

Use the Algorithm with only radio images of CHEST in DICOM format only following the HIPAA rules. Patients must have age between 1 and 95 years old.

After the Chest X-Ray is completed, the data is sent to the Algorithm. It will check the initial criteria. If it satisfies the criteria, then Algorithm will make a prediction and then send its prediction as well as the X-Ray image to a radiologist who will decide and give the final decision.

Device Limitations:

It requires the patient's position must be AP or PA.

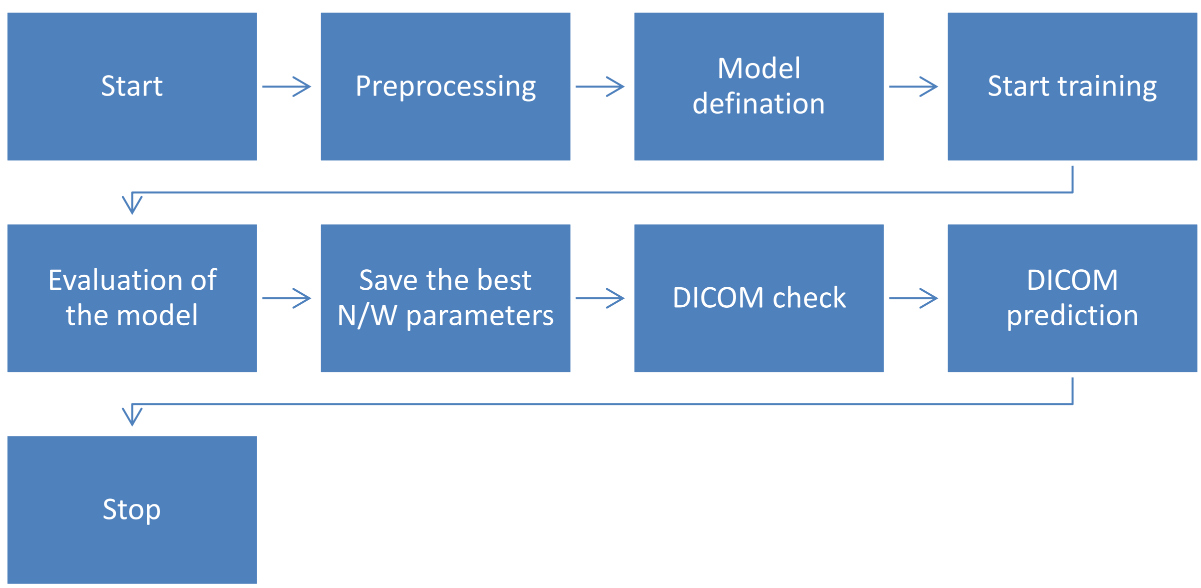
Pleural thickening and fibrosis could decrease the performance of the model because the pixel intensity distribution is quite similar to pneumonia one and the Algorithm will not be able to identify pneumonia accurately.

Clinical Impact of Performance:

There is a correlation between precision and recall. As we know If the algorithm is evaluated on precision, then, clinically, the number of correct results will surely increase and whereas the number of true positives (TP) will decrease. For optimum performance, the threshold is set to 0.44.

In the case of false-positive (FP), the Algorithm will not identify or detect existing pneumonia whereas in case of an FN (false negative) occurs the Algorithm will identify pneumonia while the patient has no pneumonia. This could significantly require the decision-making of the medical staff i.e. Radiologist.

2. Algorithm Design and Function



DICOM Checking Steps:

First DCMread and use of pixelArray. Then resizing it so that it can fit in the model. And then use DICOM to validate the model and consider the issues with the patient’s position, type of image, and recognize examined part of the body.

Preprocessing Steps:

For Normalization of DICOM image and resizing into (1,224,224,3) are required to pass the image into the Algorithm.

CNN Architecture: VGG16 with added layers.**

CNN Architecture see figure below.

```
_________________________________________________________________
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
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
fc1 (Dense)                  (None, 4096)              102764544
_________________________________________________________________
fc2 (Dense)                  (None, 4096)              16781312
_________________________________________________________________
predictions (Dense)          (None, 1000)              4097000
=================================================================
Total params: 138,357,544
Trainable params: 138,357,544
Non-trainable params: 0
_________________________________________________________________
```

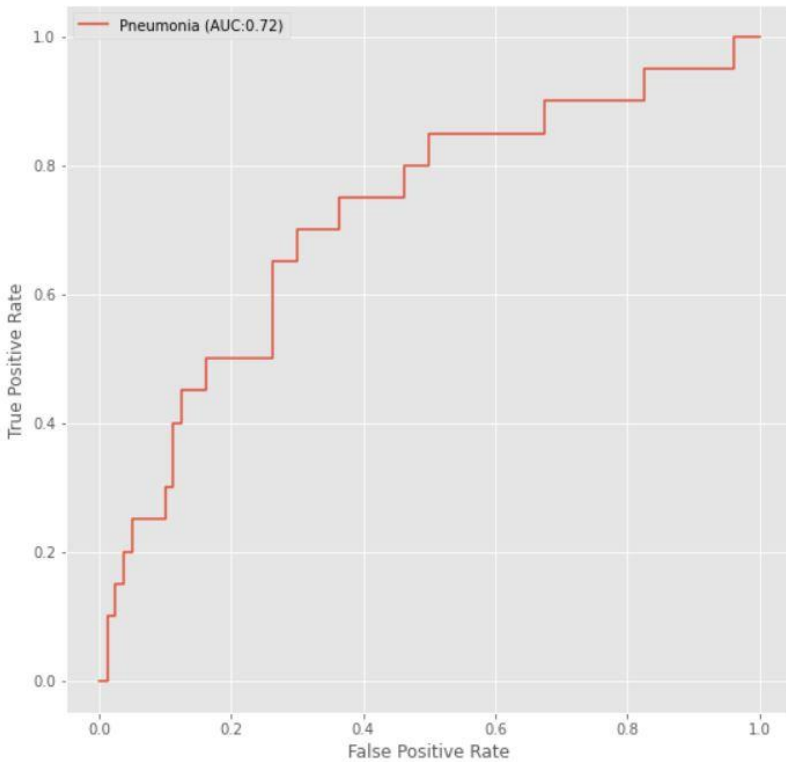
3. Algorithm Training

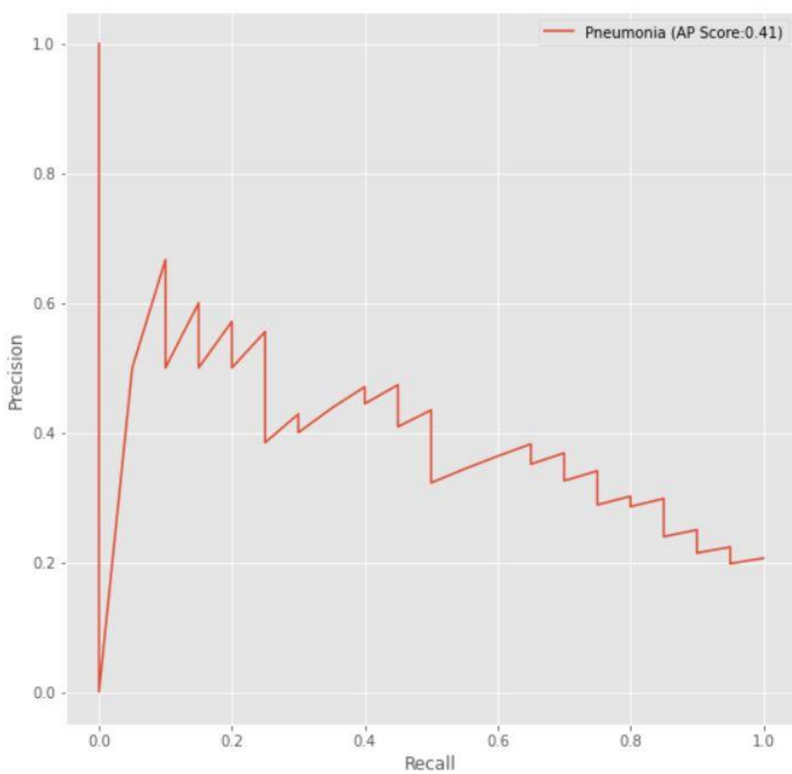
Parameters:

- Types of augmentation used during training are : rescaling, height_shift, width_shift, horizontal_flip , shear, rotation and zoom.
- Batch size: 30
- Optimizer learning rate: adam, lr = 1e-4
- Layers of pre-existing architecture that were frozen transfer: layer is "block5_pool"
- Layers of pre-existing architecture that were fine-tuned: is the output layer
- Layers added to pre-existing architecture: Flatten, Dense(1024) with relu activation, Dense(512) with relu activation, Dense(1) with sigmoid activation.

<< Algorithm training performance visualization >>

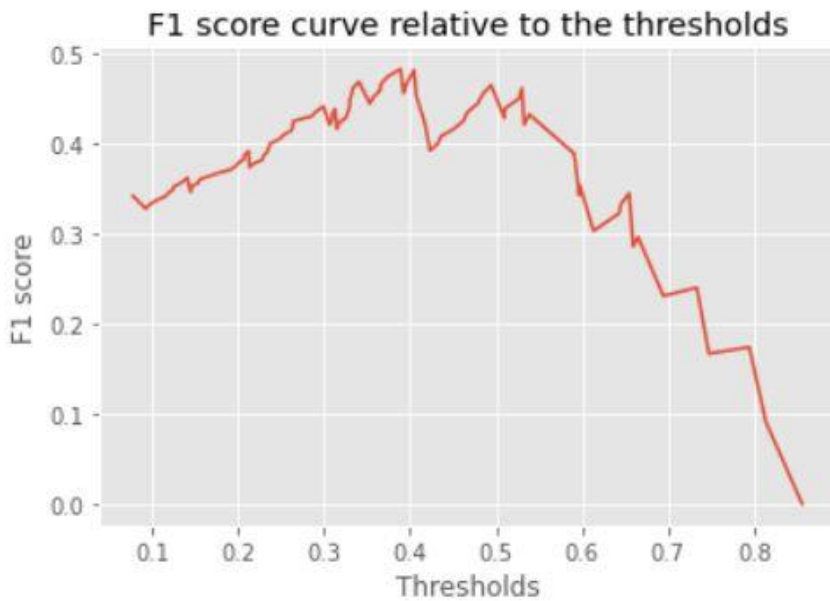
<< Insert P-R curve >>





Final Threshold and Explanation:

The threshold of 0.405 gives the best F1 score of 0.48. This means that if the prediction is above 0.44 then Algorithm identifies it and we can say that it's pneumonia. In fact, the trade-off between precision and recall is necessary as it depends on the clinical settings. High precision signifies that the number of relevant results is high, i.e. people or patients identify with pneumonia will surely have pneumonia in most cases. High recall on the other hand signifies that the number of False Negative will be low thus the number of patients having pneumonia that is identified is high.

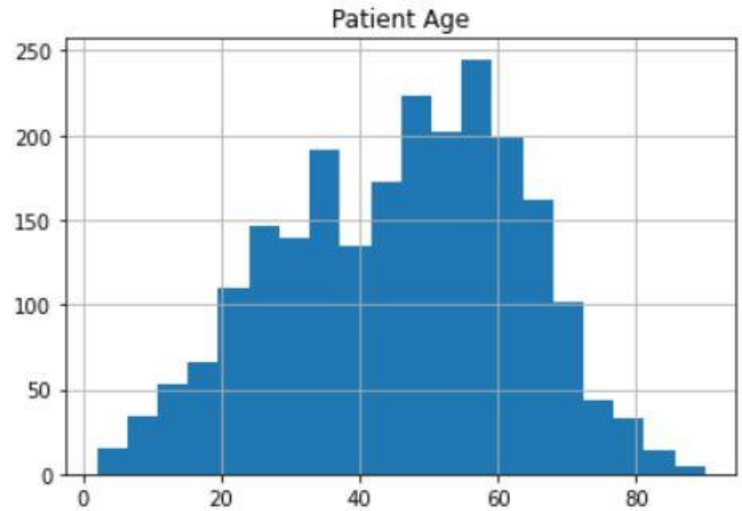


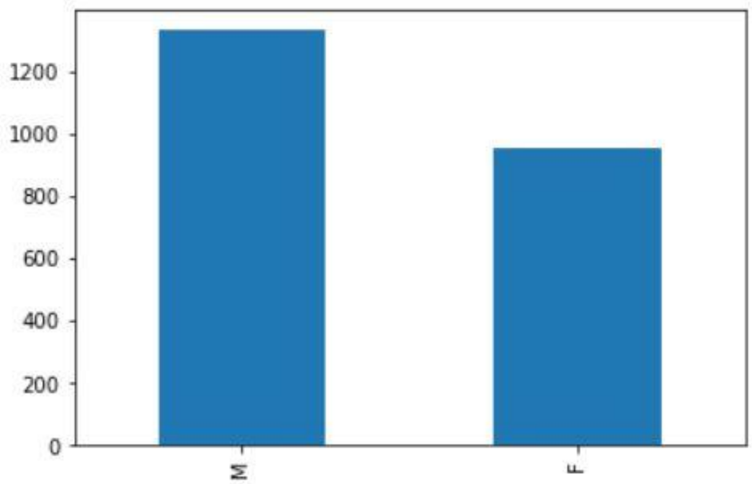
4. Databases

Description of Training Dataset:

The training dataset contains 2289 rows with information on the finding labels like follow-ups, patient ID, gender, age, images parameters such as Width and Height, and the view position. The training dataset was balanced to have an equal amount of positive and negative cases of Pneumonia.

(See below for more info, include visualizations as they are useful and relevant)

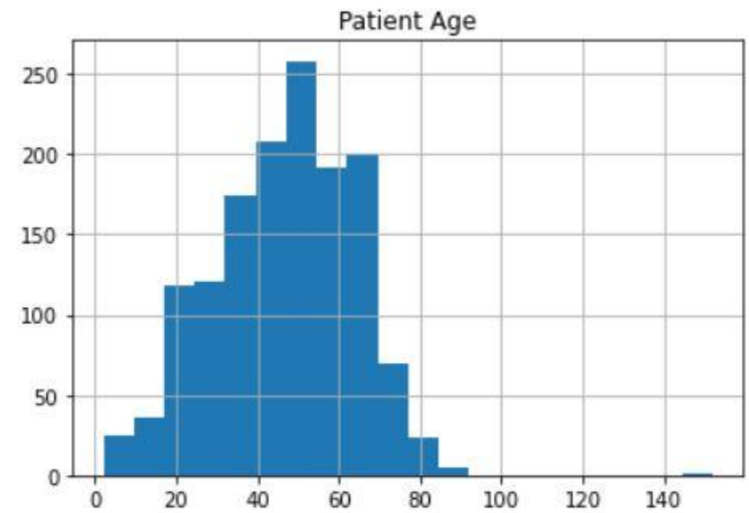


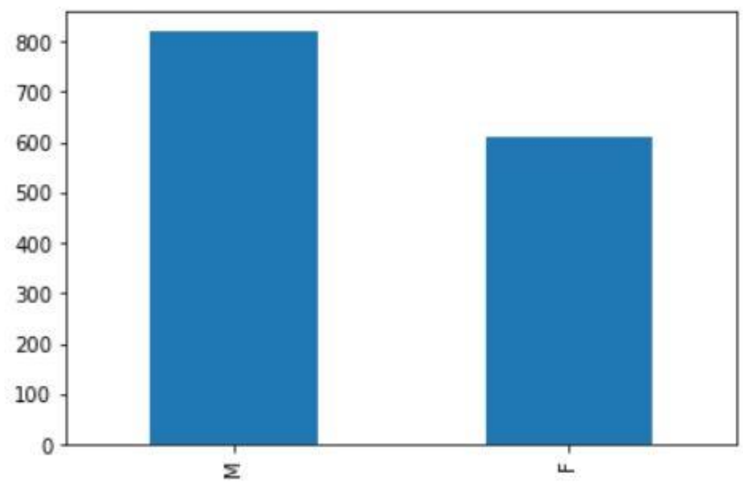


Description of Validation Dataset:

The validation dataset contains 1430 rows with information on the finding labels like follow-ups, patient ID, gender, age, images parameters such as Width and Height, and the view position. In the validation set, we balanced the dataset to have 20% of pneumonia cases to have an 80/20 classic division as we required.

(See below for more info, include visualizations as they are useful and relevant)





5. Ground Truth

To obtain the ground truth we use Natural Language Processing (NLP) to mine the associated radiological reports. The labels include 14 common thoracic pathologies such as Pneumonia, Consolidation, Infiltration, and Pneumothorax. The limitation of this dataset is that image labels were obtained by using NLP-extraction so there are possibilities of some erroneous labels but the NLP labeling accuracy is far too good i.e. it is estimated to be >90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

A dataset representing the gender and age clinical representation for the FDA validation dataset. This dataset must include a bigger number of comorbidities with pneumonia to investigate more in detail each comorbidity factor. So finally, the Algorithm worked below the level on pleural thickening and fibrosis thus a validation dataset shouldn't include these disease images.

Ground Truth Acquisition Methodology:

Ground truth can be obtained through biopsy labeling, a biopsy is the ultimate gold standard, and to check have gold standard ground truth. However, due to time and cost constraints, a silver standard acquired through a voting system of radiologists could be a proven sufficient standard.

Algorithm Performance Standard:

CheXNet on the same provided dataset obtained an F1 score of 0.435. For training the model, the metric is used as the binary cross-entropy loss. This algorithm obtained an F1 score of 0.48 which is quite better than the performance score of the radiologists with an average of 0.38.