

# FDA Submission

---

**Your Name:** Yassine KHIARI

**Name of your Device:** PneumoDetectNet

## Algorithm Description

### 1. General Information

**Intended Use Statement:** For assisting a radiologist in detection of pneumonia in x-ray images.

**Indications for Use:**

1. Screening of x-ray images.
2. The age of patient should be between 2 and 90 as indicated in the age distribution of data used to train our model.
3. Can be used for both men and women as distribution of the gender is quite balanced.
4. Body part of the X-Ray image must be "chest"
5. Body Position must be AP or PA
6. Modality should be DX

**Device Limitations:** As observed in the EDA file the Pneumonia coexists with several diseases.

However the most frequent one like Infiltration, Edema and Atelectasis might confuse our model prediction due to the fact that it won't be able to differentiate between the diseases. Hence it is recommended to do not use the model when one of the listed above disease is present.

**Clinical Impact of Performance:**

From the observations in the EDA, we can see the trade-off precision recall. We had to make a choice in order to design this model to be either high precise or have high recall.

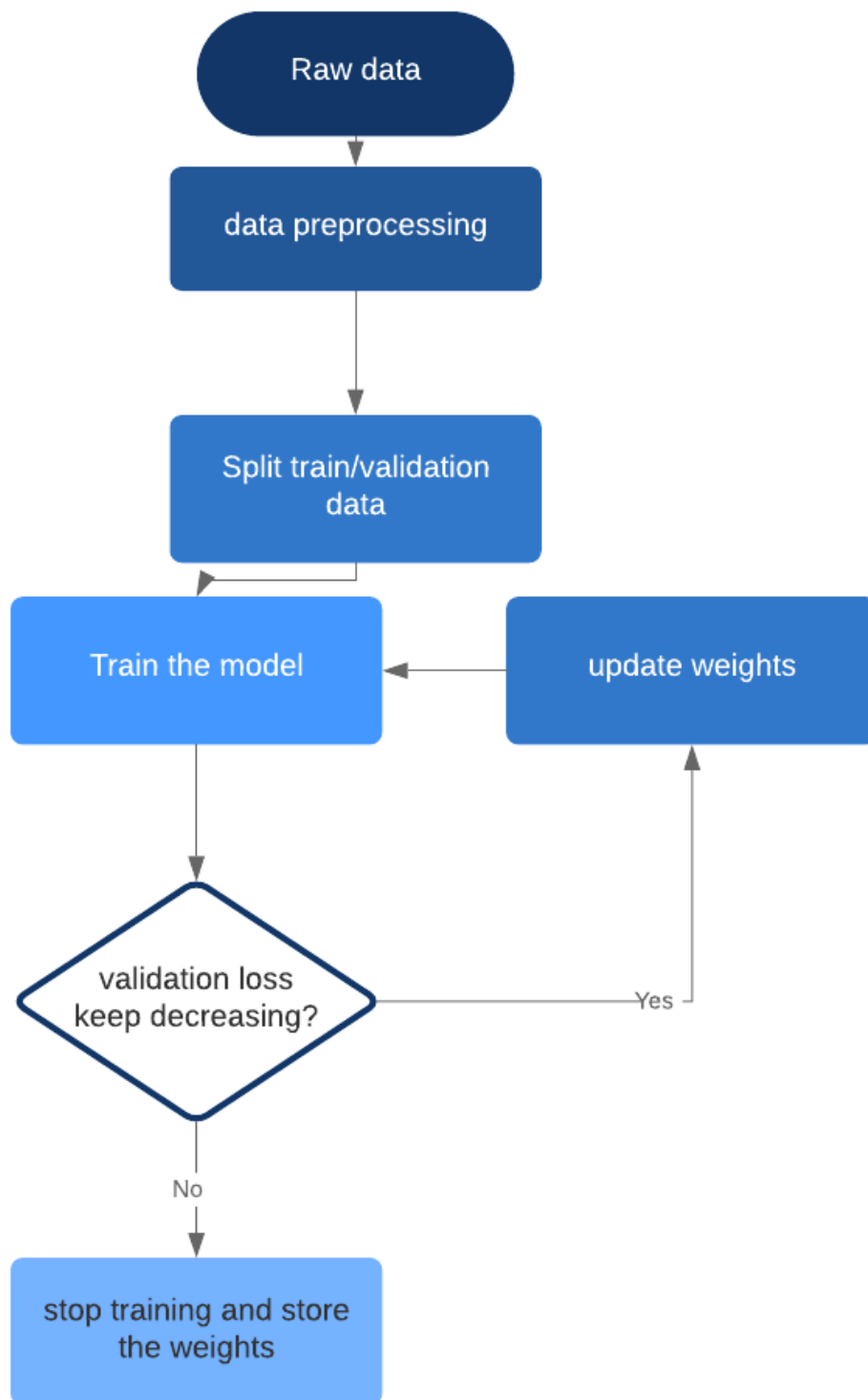
This model could have been more oriented to precision such that the rate of False Positive is very low.

or

This model could have been more oriented to recall such that the rate of False Negative is very low.

In the end improving precision typically reduces recall and vice versa so as we are using this model to assist a real radiologist we chose the best threshold from the F1 curve and finally it should be enough to screen the patients.

### 2. Algorithm Design and Function

**DICOM Checking Steps:**

The algorithm performs the following checks on the DICOM image:

- Check Patient Age is between 2 and 90 (inclusive)
- Check Examined Body Part is 'CHEST'
- Check Patient Position is either 'PA' (Posterior/Anterior) or 'AP' (Anterior/Posterior)
- Check Modality is 'DX' (Digital Radiography)

**Preprocessing Steps:**

The algorithm performs the following preprocessing steps on an image data:

- Converts RGB to Grayscale (if needed)
- Re-sizes the image to 244 x 244 (as required by the CNN)
- Normalizes the intensity to be between 0 and 1 (from original range of 0 to 255)

**CNN Architecture:** CNN Architecture: The base network is VGG16 pre-trained on ImageNet dataset, followed by:

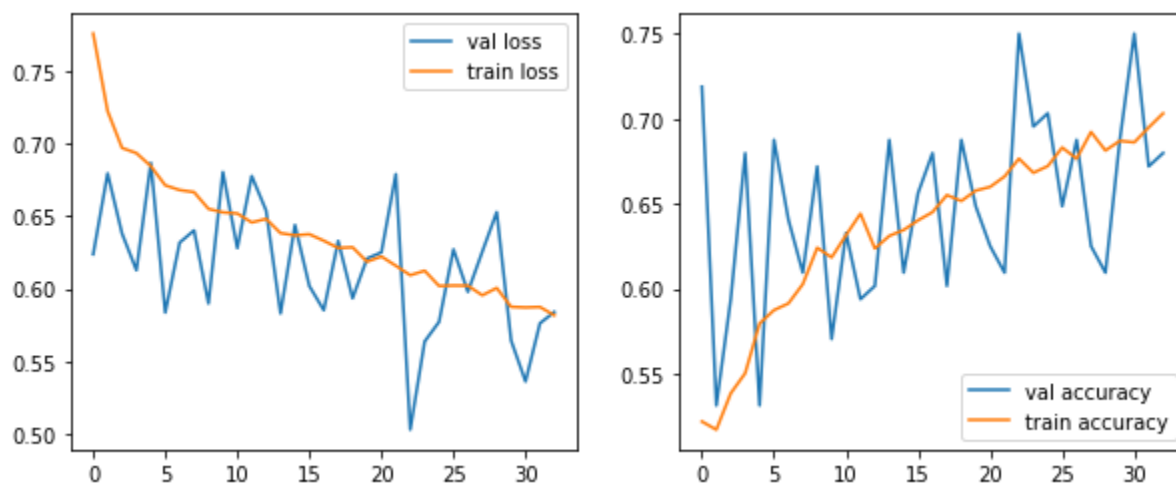
- Flatten()
- Dropout(0.5)
- Dense(1024, activation='relu')
- Dropout(0.5)
- Dense(512, activation='relu')
- Dropout(0.5)
- Dense(256, activation='relu')
- Dense(1, activation='sigmoid')

### 3. Algorithm Training

#### Parameters:

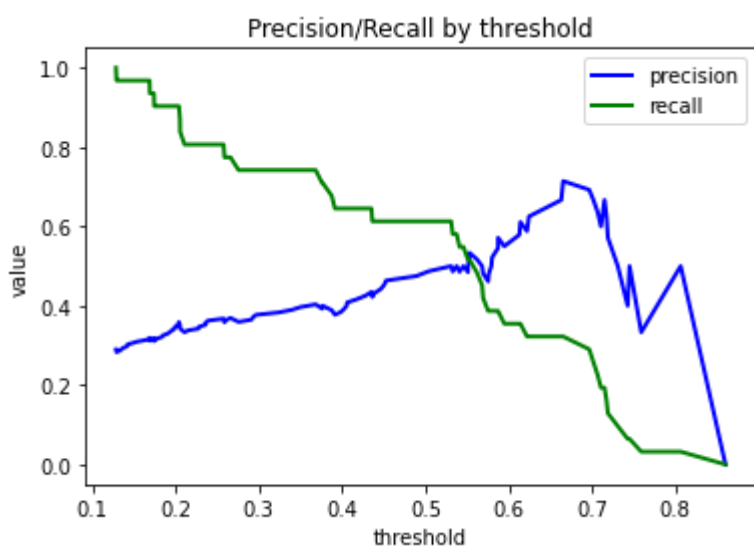
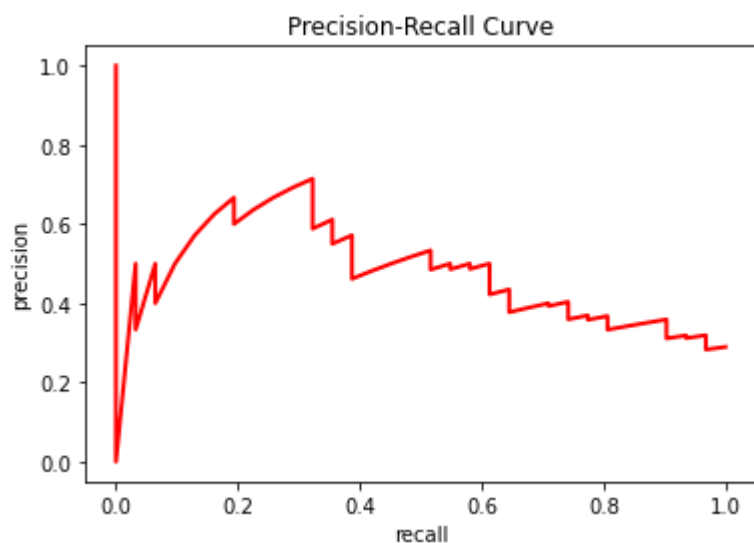
- Types of augmentation used during training
  - Horizontal flip
  - Random height shift of (+/-)10% of the image max.
  - Random width shift of (+/-)10% of the image max.
  - Random rotation shift of (+/-)10 degrees max.
  - Random shear shift of (+/-)20% max.
  - Random zoom of (+/-)15% max.
- Batch size
  - 128
- Optimizer learning rate
  - 1e-4
- Layers of pre-existing architecture that were frozen
  - First 17 layers of VGG16
- Layers of pre-existing architecture that were fine-tuned
  - All dense layers of VGG16
- Layers added to pre-existing architecture
  - Flatten()
  - Dropout(0.5)
  - Dense(1024, activation='relu')
  - Dropout(0.5)
  - Dense(512, activation='relu')
  - Dropout(0.5)
  - Dense(256, activation='relu')
  - Dense(1, activation='sigmoid')

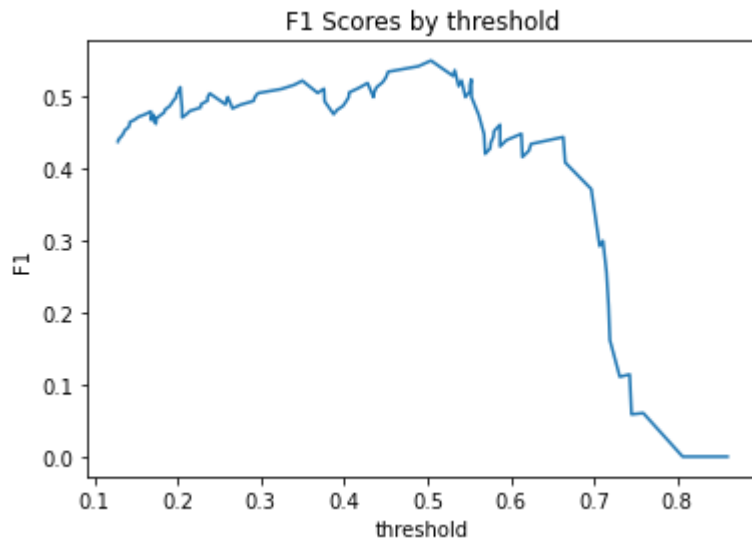
#### Algorithm training performance visualization



This Training/validation plots shows how the training of the model went through time. The high fluctuation of the validation curve shows that it is better to try retraining the model using a lower learning rate.

### P-R curves





### Final Threshold and Explanation:

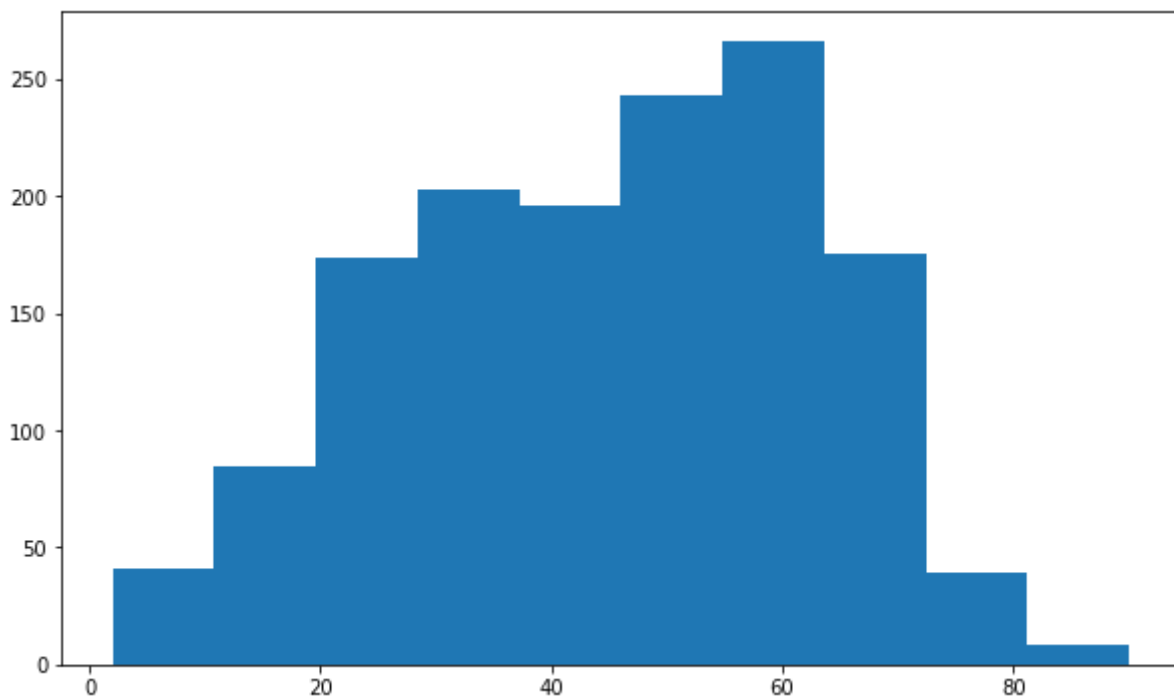
The threshold 0.55 was finally choosed according to what is seen in the two previous curves. it can be clearly seen from the second curve (precision-recall) that the two best values of PR is when they meet exactly around [0.45,0.6].

However after plotting the F1 score curve it became obvious that the best threshold value for this model is 0.55.

## 4. Databases

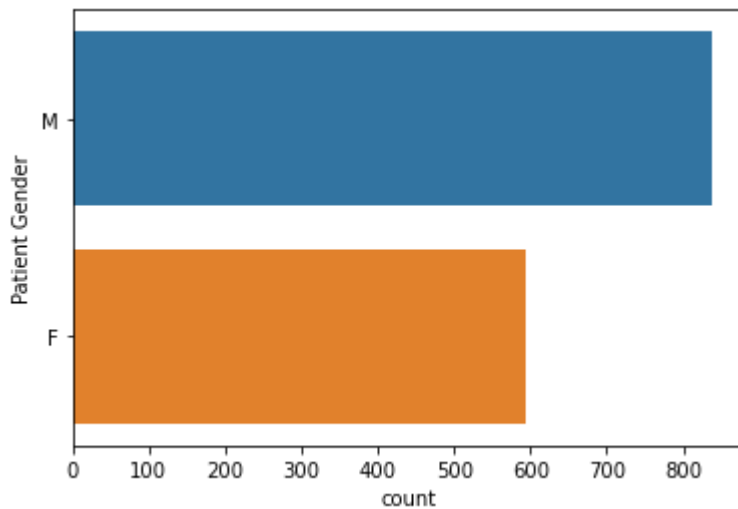
The following chart describe the age distribution of Pneumonia patients:

- We can clearly see that the age is distributed between 2 and 90.
- This dataset was used for train and validation set.



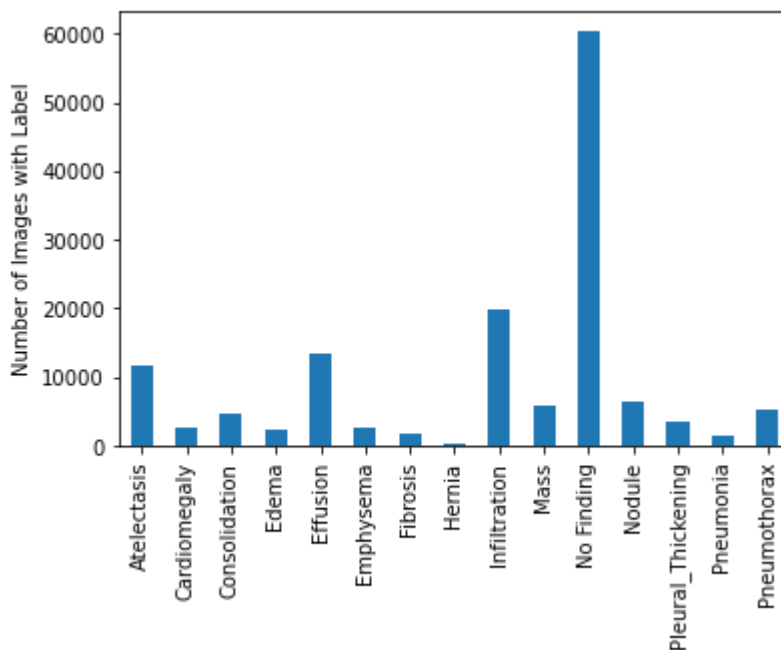
The following chart describe the sex distribution of Pneumonia patients:

- We can clearly see that the gender is fairly balanced between male and female with a small preference to male.
- This dataset was used for train and validation set.



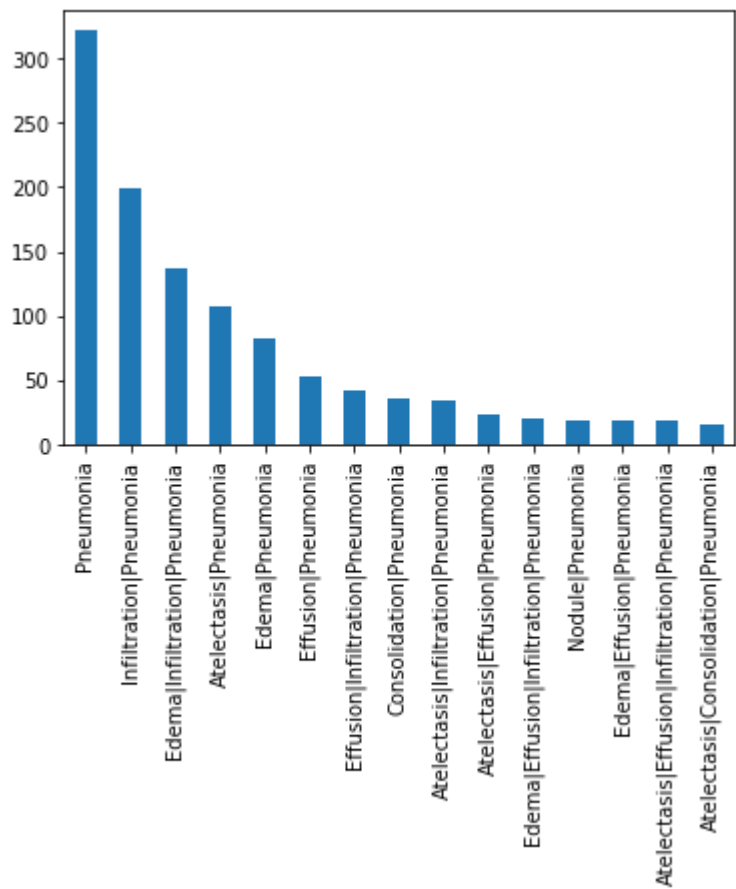
The following chart describe the number of images per label in our overall dataset:

- We can clearly see that "No Findings" has the most % of images
- Pneumonia is only represented by 1.2%.
- Hence the choice to perform data augmentation.



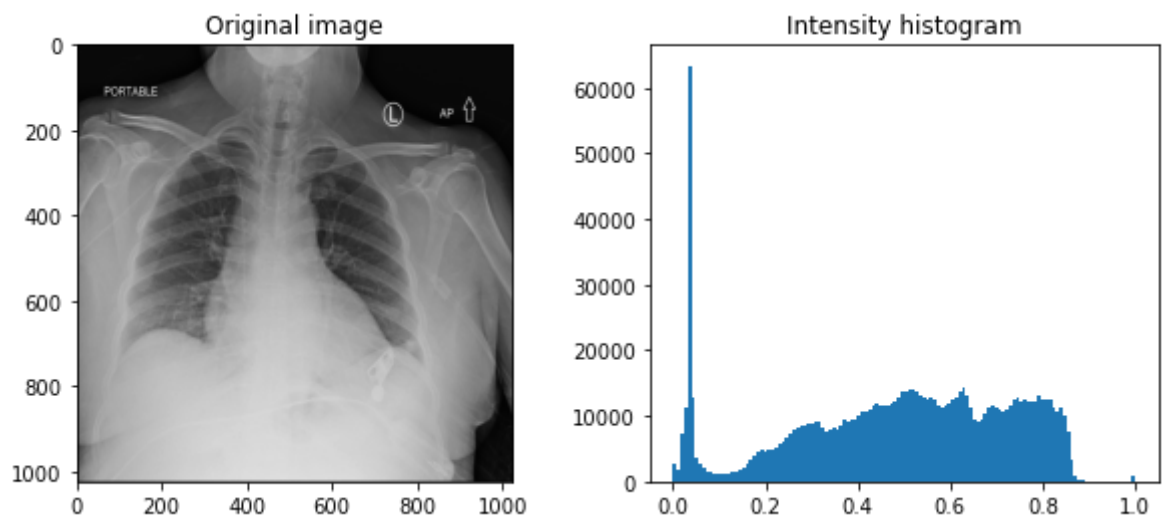
The following chart describe the coexistence of diseases in our overall dataset:

- The pneumonia disease can coexist with other diseases.
- The three mains one are:
  - Pneumonia + Infiltration.
  - Pneumonia + Edema.
  - Pneumonia + Atelectasis



The following chart describe the intensity level of a random Pneumonia Patient:

- after analyzing the dataset it has been seen that this is chart is quiet representative for the rest of the patients
- The pixel intensity level is bimodal.



Finally the split between train and test data has performed as following:

- Perfectly balanced for training data
- 20 P / 80 N % split for the validation dataset

**Description of Training Dataset:**

The training Dataset was constructed from the original provided dataset NIH Chest X-rays which consists on:

Over 112,000 Chest X-ray images from more than 30,000 unique patients.

Training dataset consists of 2290 chest xray images, with a perfect balanced labels between positive and negative cases.

### **Description of Validation Dataset:**

The validation Dataset was constructed from the original provided dataset NIH Chest X-rays which consists on:

Over 112,000 Chest X-ray images from more than 30,000 unique patients.

Training dataset consists of 1430 chest xray images, with an unbalanced labels, 20% positive and 80% negative cases.

## **5. Ground Truth**

The labels are obtained using an NLP approach from the radiologist reports.

Hence, using this method, it is highly expected to have some wrong labels.

Besides the Patient's history were not provided so the labels were not based on the history which might impact the label of the image.

## **6. FDA Validation Plan**

### **Patient Population Description for FDA Validation Dataset:**

- Gender: Both men and women
- Age: Between 2 to 90
- Known diseases: Without known the diseases (Infiltration, Edema, Atelectasis)
- Body part: Chest
- imaging modality: DX
- Patient position: 'PA' or 'AP'

### **Ground Truth Acquisition Methodology:**

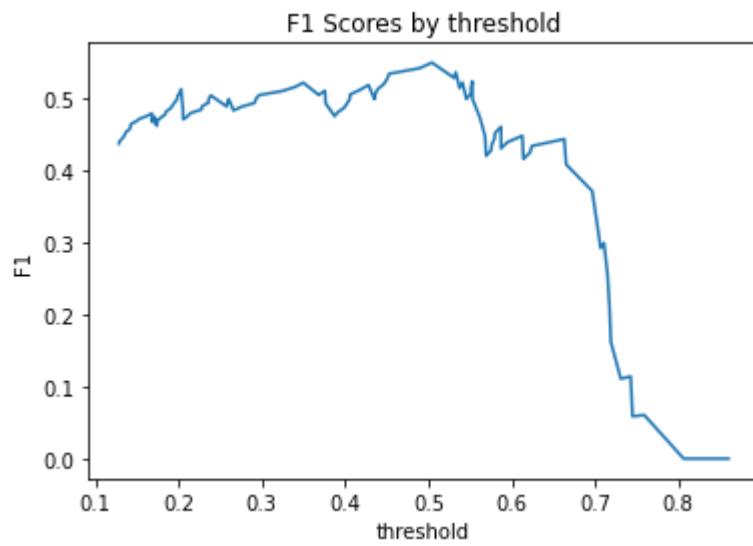
Since the aim of this model is only to help a radiologist screening patient we can use the 'silver standard' as ground truth which is average score of three radiologist.

### **Algorithm Performance Standard:**

The model should be tested on f1-score and the minimum acceptable f1-score should be 0.387 bas stated in this paper. <https://arxiv.org/pdf/1711.05225.pdf> According to all that and



based on the chart here:



The optimal F1 value computed for our model is 0.5.