FDA submission

Humberto Jimenez Martinez

Pneumonia detection assistant

Algorithm Description

1. General Information

Intended use statement:

This tool is meant to assist Radiologists with the detection and diagnosis of Pneumonia in patients by analyzing their x-ray scans.

Indications for use:

The tool works by reading x-ray scan images in DICOM format. The images need to be chest scans with Posterior Anterior (PA) and Anterior Posterior (AP) views. The algorithm will indicate the probability of presence of pneumonia.

Device limitations:

The algorithm presents a high recall rate, meaning that it will reduce the likelihood of false negatives. The precision rate is likely to mistakenly identify other conditions as Pneumonia.

Clinical impact of performance:

The tool was designed to assist the radiologists, it should not be taken as a diagnosis tool. Its role is to validate and verify the professional opinion of a clinical expert.

The advantage of the algorithm is the high recall rate which prevents the likelihood of false negatives, which when undetected has a major impact on patients wrongly discharged from the medical care.

2. Algorithm Design and function

The following chart depicts the process of design and creation of the algorithm.

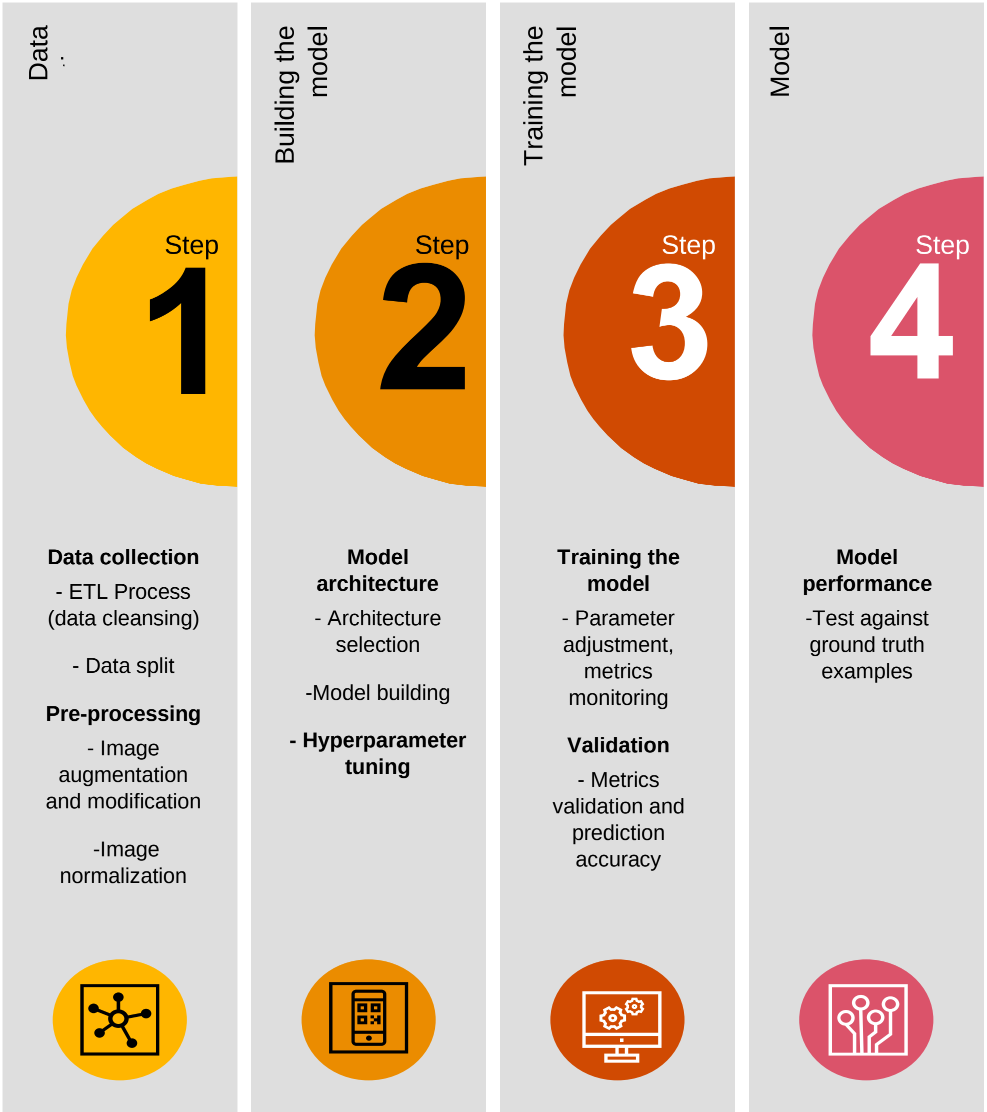

Figure 1 – Algorithm Flowchart

DICOM checking steps:

As previously mentioned, the DICOM images need to be classified as 'CHEST' x-rays in either Posterior Anterior (PA) and Anterior Posterior (AP) views. The modality of the image needs to be 'DX'.

Pre-processing steps:

All images were resized to a (224,224, 3) format

All images were normalized to reduce the mean and the diminish the standard deviation to nil.

CNN architecture:

The model was built by using a pre-trained VGG16 model as the base model. The fully connected layers were replaced with new layers as follows:

- Flatten layer
- Dropout (0.5)
- Dense layer (1024, relu activation)
- Dropout (0.5)
- Dense layer (1024, relu activation)
- Dropout (0.5)
- Dense fc layer (1, sigmoid activation)

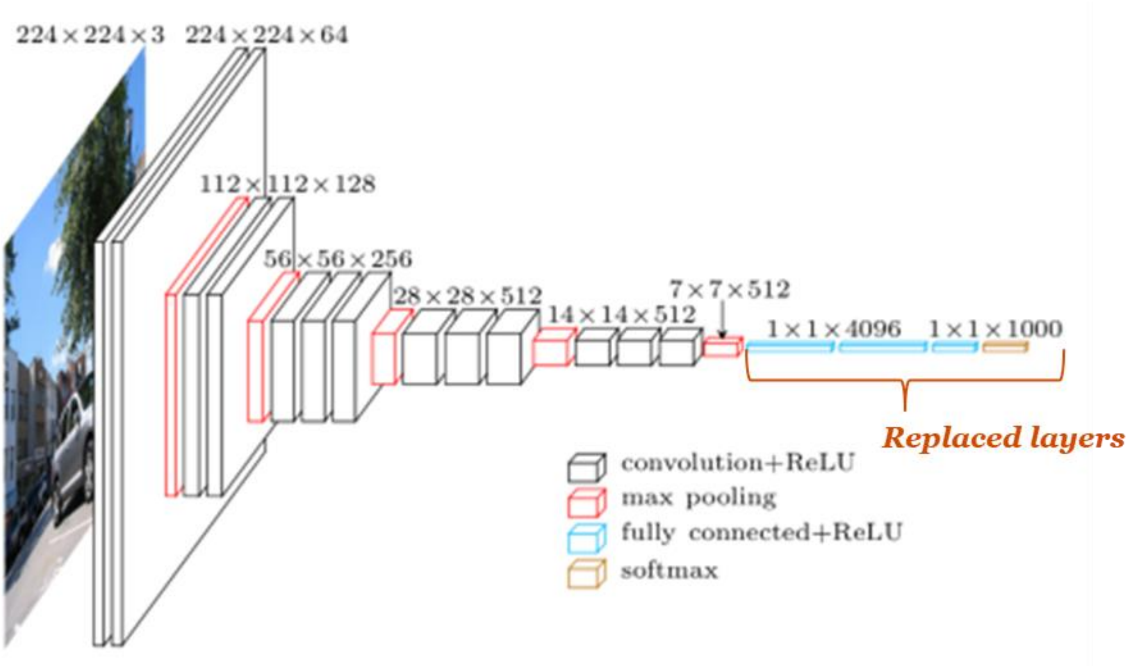

Figure 2 - CNN Architecture example

3. Algorithm Training

Parameters:

Type of augmentation:

The pre-processing step allowed us to train the model with a rich variety of images. For the training images the following steps were taken:

- Rescaling the images by dividing them by a factor of 255
- Horizontal flip of images
- Width shift of 0.1
- Rotation range of 10
- Shear range of 0.05
- Zoom range of 0.05

Validation images were not augmented, they were only normalized.

Batch size: 16 images per batch

Optimizer learning rate: starting at 0.0001

Layers of pre-existing architecture that were frozen: 10 convolutional layers and 4 maxpool layers

Layers of pre-existing architecture that were fine-tuned: 3 convolutional layers and one maxpool layer

Layers added to pre-existing architecture: 1 flatten layer, 3 dense layers and 3 dropout layers.

Model metrics:

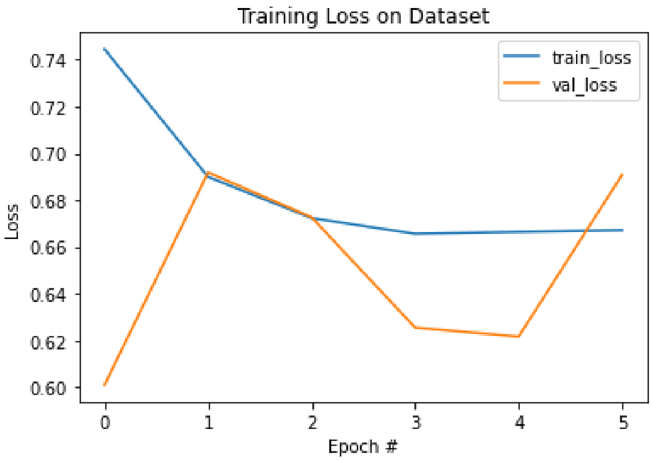

Figure 3 - Model training history

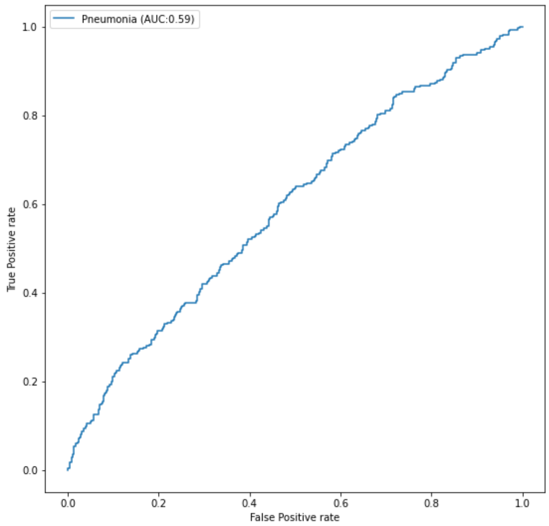

Figure 4 – ROC Curve

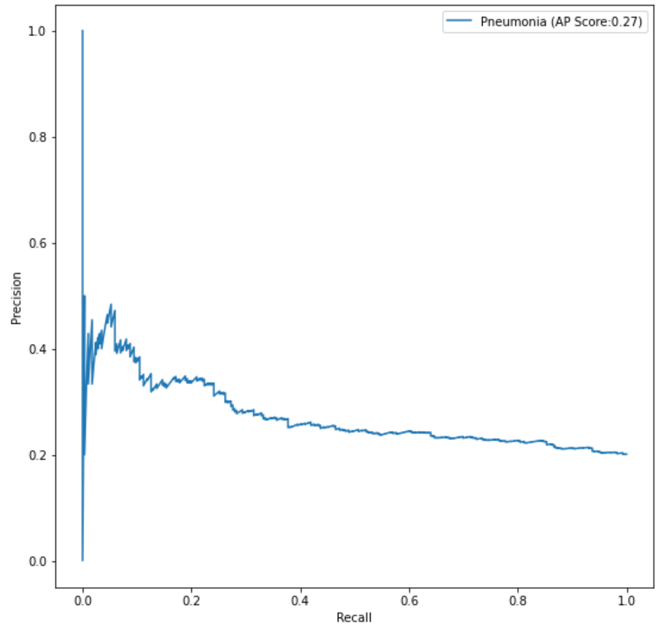

Figure 5 - Precision-Recall curve

Final threshold and explanation:

The model showed poor precision regardless of the threshold, therefore the decision to pursue recall was a better usage of the model. The final threshold selected was 0.35 which gives a recall of around 0.8. The precision for that threshold is 0.22 and the F1 score is 0.35.

4. Databases

Description of the training dataset:

The dataset contained 112,120 x-ray images, in which 1,430 were Pneumonia positive cases (1.28%) of the total. A little over 55% were images with no findings, the rest of the images had 1 or multiple conditions ranging from 14 types (Pneumonia inclusive)

A few outliers where the patients age was above 100 years old were removed from the data set to ensure data integrity and accuracy.

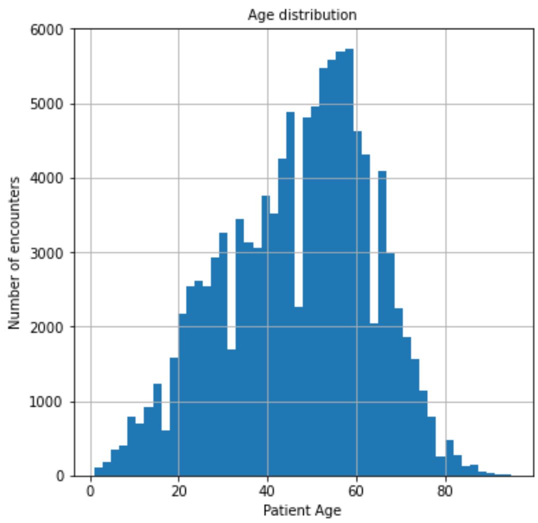

Figure 6 -Age distribution

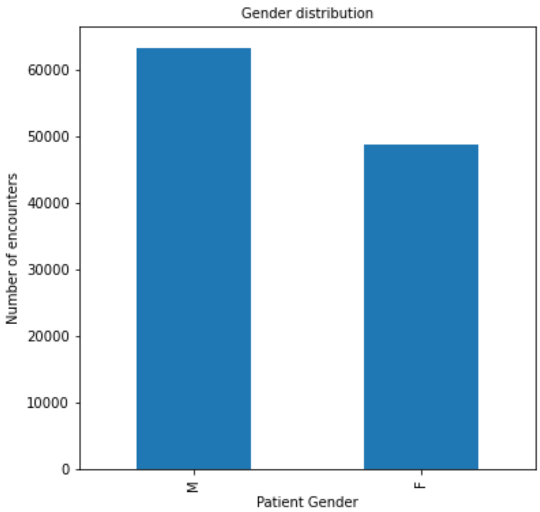

Figure 7 - Gender distribution

In terms of age distribution, the majority of the patients had a median age of 50 years old, the range presents a normal distribution of values. In terms of gender, 56% of the patients are biologically identifies as males, while 44% are identified as females.

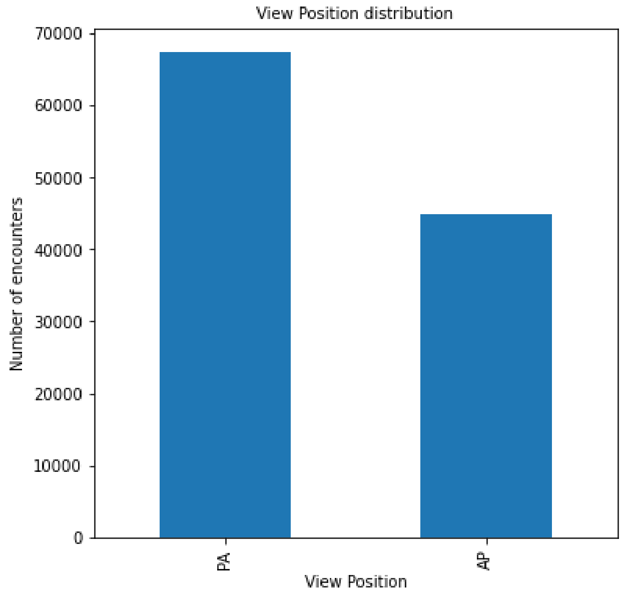

Figure 8 - View position distribution

In terms of image view position, 60% of the images are Posterior Anterior (PA) and 40% Anterior Posterior (AP).

Description of the validation dataset:

The validation dataset consisted of 1,430 in which 20% were Pneumonia positive, the split of images assured that there was enough representation of the condition in both split datasets (training and validation)

5. Ground Truth

The ground truth used for this model were the labels from the dataset. These labels were extracted using NLP, there is the possibility that these labels are not 100% accurate. This limitation is acknowledge when building the model, but there is confidence that the accuracy is not below 90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

The patient population considered for validation should be between the age range of 1- 100 years old, gender ratio 1:1 Males and females, with or without previous history of lung conditions.

Ground Truth Acquisition Methodology:

All images should be captured in either Posterior Anterior (PA) or Anterior Posterior (AP), expert diagnosis is required and the existence of other conditions including: Cardiomegaly, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Atelectasis, Pleural thickening, Consolidation, Nodules, Mass, and Pneumothorax, should be accurately detected and labelled.

Algorithm Performance Standard:

Using a panel of trained radiologists is recommended to achieve adequate performance.