FDA Submission

Author: Francisco Londono

Name of the Device: Pneumonia assisting model

Algorithm Description

1. General Information

- Intended Use Statement:
  For assisting the radiologist to detect pulmonary abnormalities compatible with pneumonia on chest x-rays.

- Indications for Use:
  This algorithm is intended for use on male and female patients from the ages 18-65 years with an indication of a chest frontal-view X-ray image in posterior-anterior (PA) and anterior-posterior (AP) position for screening pulmonary pathologies.

  In the clinical workflow, x-ray images should be delivered in DICOM format to the algorithm, which checks basic conditions, such as modality and body part or view of the patient, to make a prediction. This prediction is then used by the expert radiologist/doctor, together with other examination results, to give a diagnosis and decide treatment.

- Device Limitations:
  The device should be used as an assistance tool for an expert, who ultimately will give the final diagnosis based in complementary results.

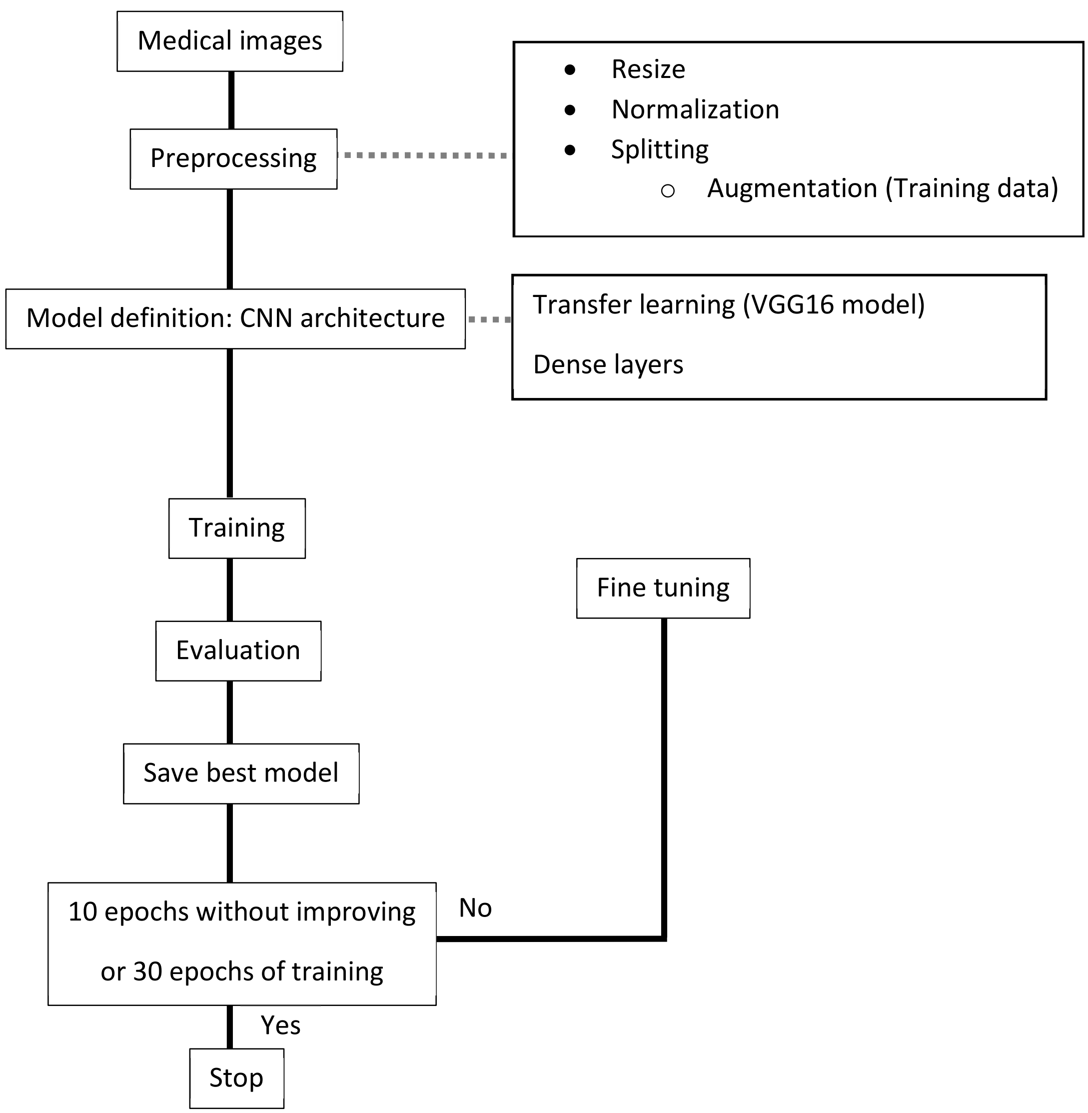
  Even if comorbidities of pneumonia, such as infiltration, edema or atelectasis, could impact negatively the performance of the algorithm, it would be important the advise of an expert to determine if these comorbidities could be a sign of pneumonia itself and should be included in the analysis and training of the algorithm.

  The model can run in a device without any special computational tools, such a GPU, to perform a quick analysis and deliver predictions.

- Clinical Impact of Performance:
  As the algorithm results are complementary to other diagnostic tools in the clinical flow for pulmonary pneumonia detection, the impact of false negatives or false positives leads to the necessity of additional screening and support diagnosis to confirm results.

  A false positive implies the confirmation of the diagnosis and additional examination for the discomfort and stress of the patient.
  However, a false negative can impede the adequate treatment of a serious pulmonary condition that can ultimately results in the aggravation or even death of the patient.

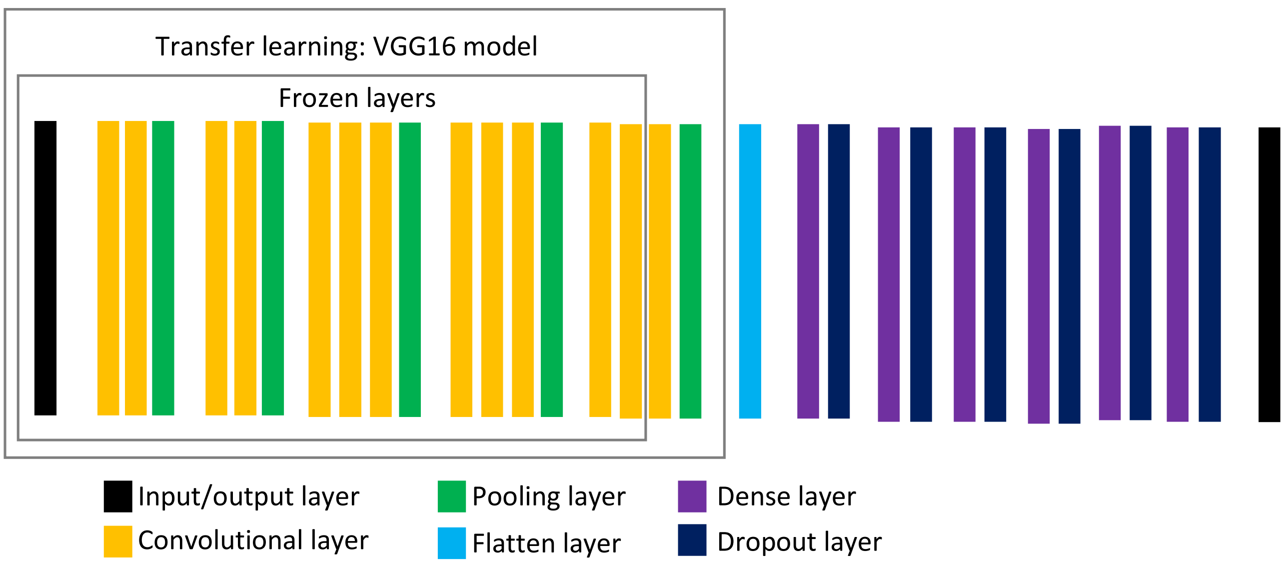
2. Algorithm Design and Function



- DICOM Checking Steps:
  - Modality: x-rays (DX)
  - Body part examined: chest
  - Patient position: Anterior-posterior (AP) or posterior-anterior (PA)

- Preprocessing Steps:
  - Normalization
  - Resize

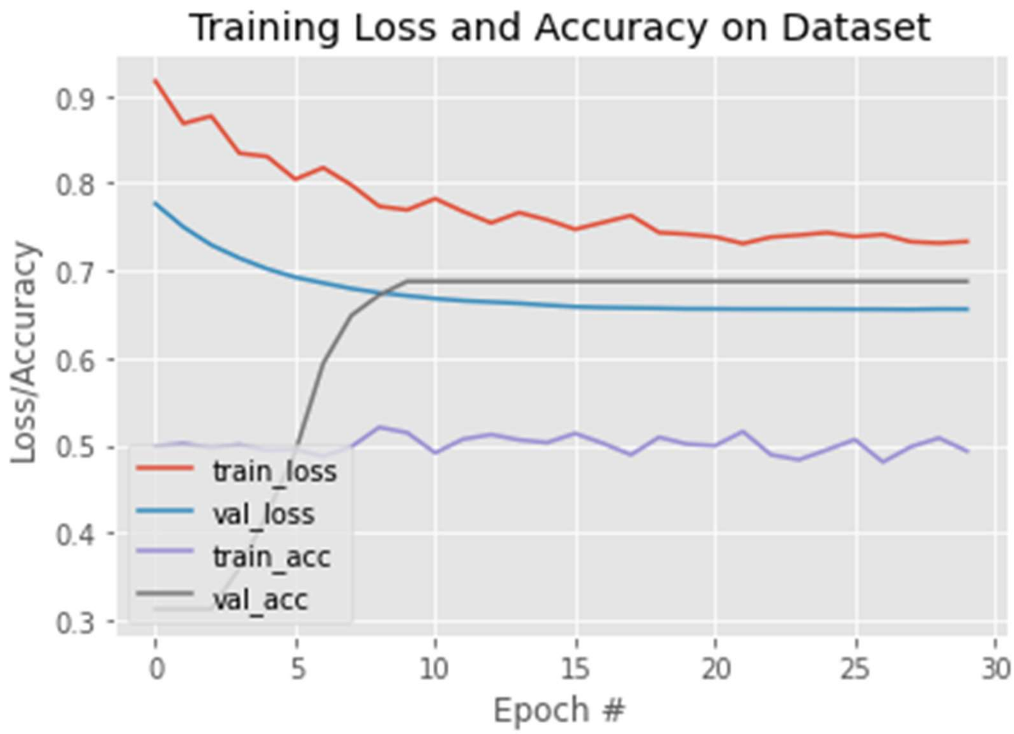
- CNN Architecture:



3. Algorithm Training

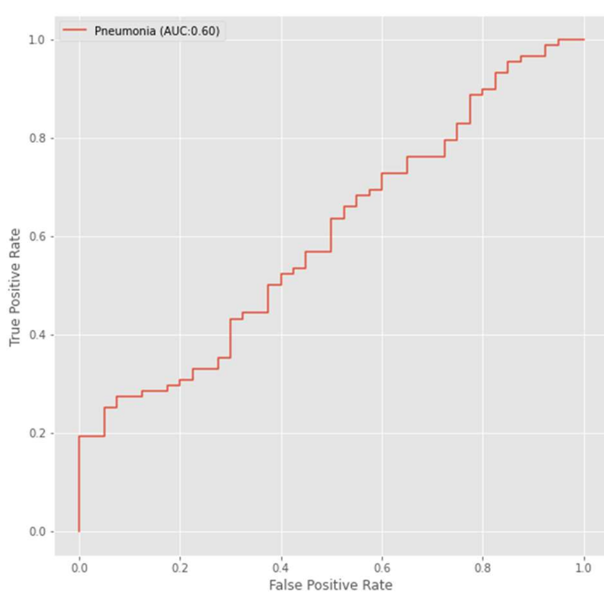
- Parameters:
  - Types of augmentation used during training:
    - rescale=1. / 255.0,
    - horizontal_flip = True,
    - vertical_flip = False,
    - height_shift_range= 0.1,
    - width_shift_range=0.1,
    - rotation_range=15,
    - shear_range = 0.1,
    - zoom_range=0.1,

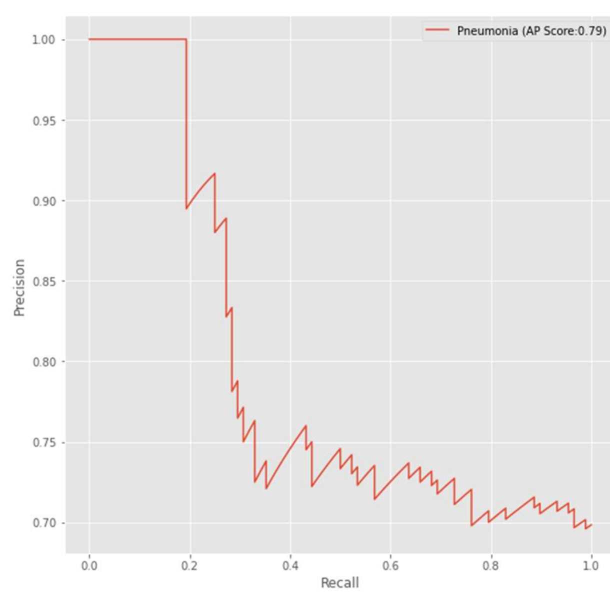
- Batch size: 256
- Optimizer learning rate: Adam at a learning rate of 1e-6
- Layers of pre-existing architecture that were frozen (VGG16 architecture):
    - input_1
    - block1_conv1
    - block1_conv2
    - block1_pool
    - block2_conv1
    - block2_conv2
    - block2_pool
    - block3_conv1
    - block3_conv2
    - block3_conv3
    - block3_pool
    - block4_conv1
    - block4_conv2
    - block4_conv3
    - block4_pool
    - block5_conv1
    - block5_conv2
- Layers of pre-existing architecture that were fine-tuned
    - block5_conv3
    - block5_pool
- Layers added to pre-existing architecture
    - Flatten layer
    - Dense1 layer ("relu" activation)
    - Dropout1 layer
    - Dense2 layer ("relu" activation)
    - Dropout2 layer
    - Dense3 layer ("relu" activation)
    - Dropout3 layer
    - Dense4 layer ("relu" activation)
    - Dropout4 layer
    - Dense5 layer ("relu" activation)
    - Dropout5 layer
    - Dense6 layer ("relu" activation)
    - Dropout6 layer
    - Output layer (Dense "sigmoid" activation)



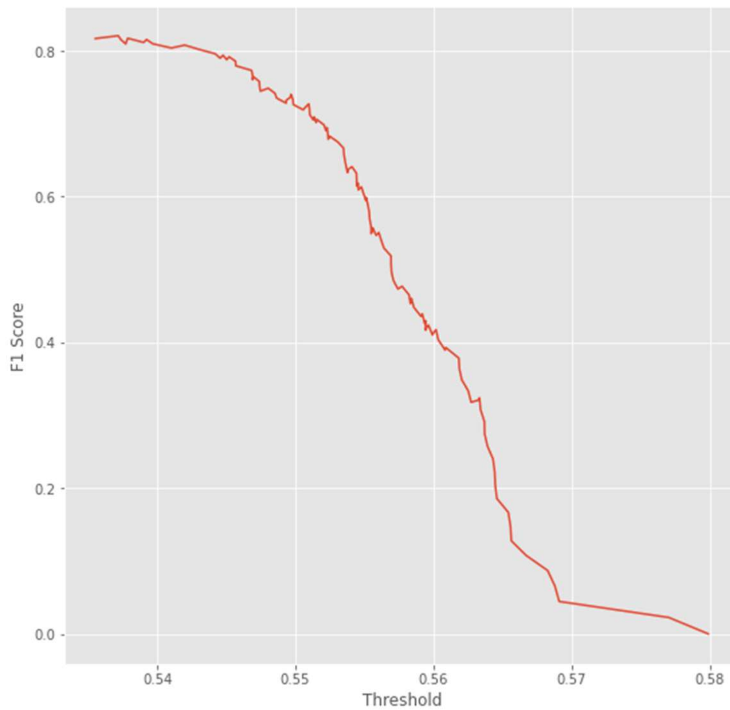
Training and validation loss decreased during the first 15 epochs and, after, validation loss stabilized while training loss continue to show a slight decrease trend.

Validation accuracy improved during the first 10 epochs and then stabilized, while training accuracy was stable during all epochs.





- Final Threshold and Explanation:



Precision is: 0.696

Recall is: 0.9886363636363636

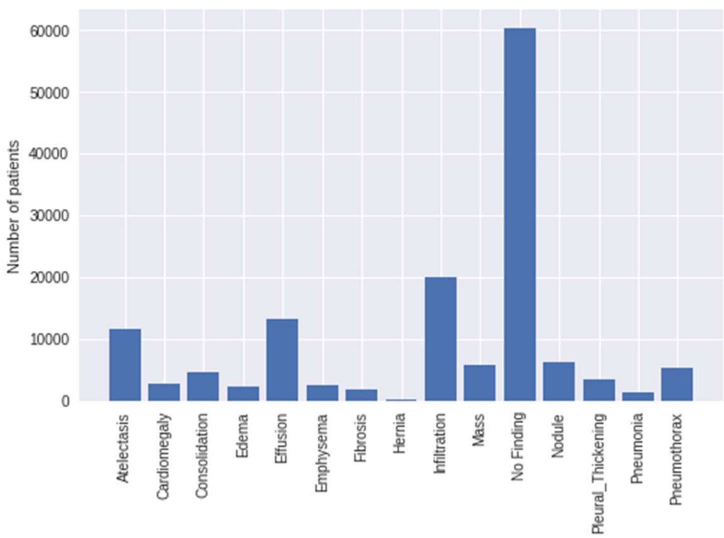
Threshold is: 0.5371808

F1 Score is: 0.8169014084507042

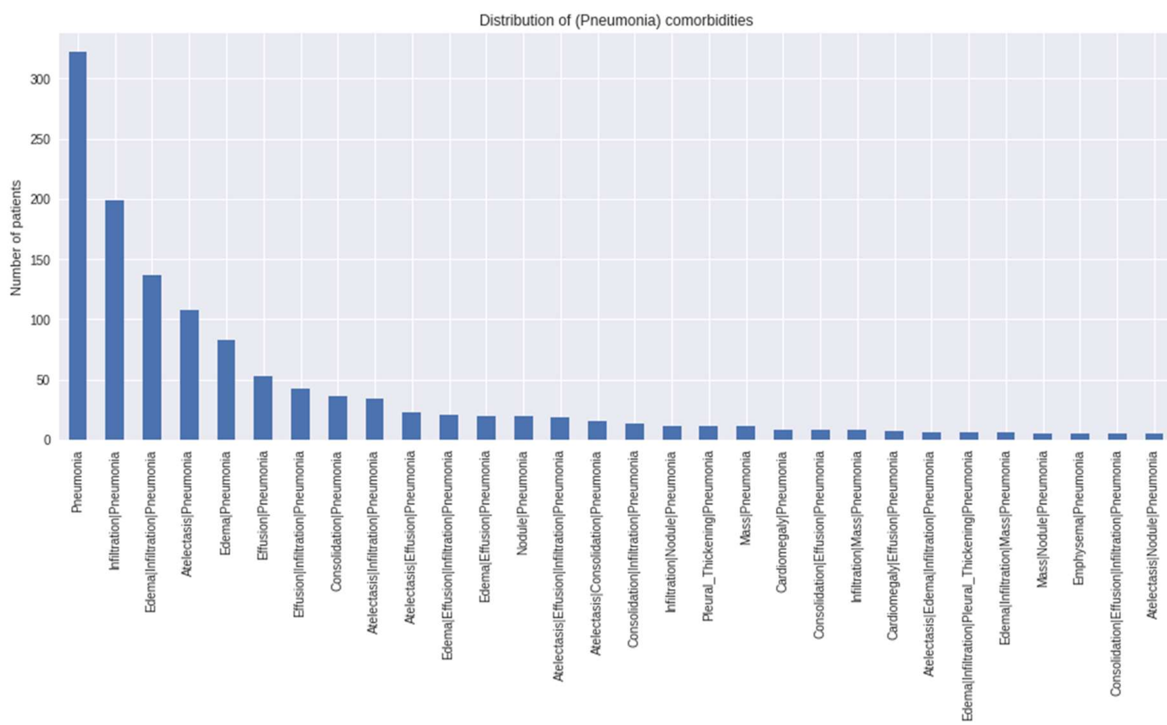
This threshold determined the best F1 score for the algorithm performance.

4. Databases

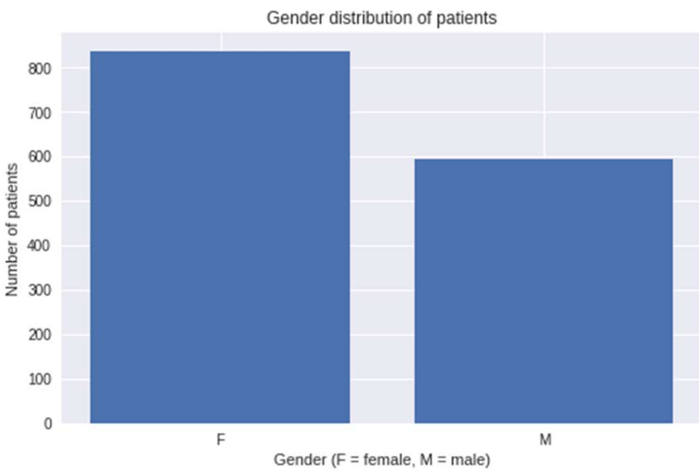
Most images in this dataset are classified as negative for pulmonary diseases ("No Findings"), whilst the most common diagnoses are Infiltration, Atelectasis, Effusion. Pneumonia is one of the least common labels, and its age distribution shows that most cases ranged between 18 and 65 years old.

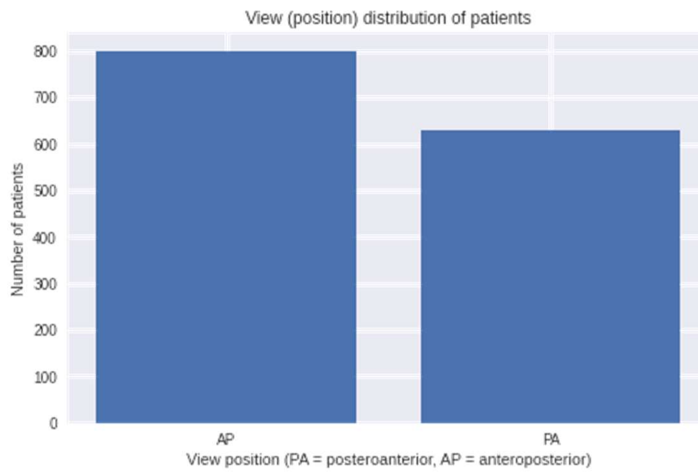


The most common comorbidities for pneumonia are Infiltration, Atelectasis, Edema and Effusion.



In the pneumonia dataset, female patients are slightly more prevalent than male cases, and more cases have an AP (anterior-posterior) view than an PA (posterior-anterior) view.





To train and evaluate the model, the dataset was split using a stratification based on the pneumonia cases distribution, and selecting 20% of the data for validation.

- Description of Training Dataset:

  Further selection was necessary to balance the training dataset. At the end, 2290 chest x-rays images were selected, in which the amount of positive and negative pneumonia cases was balanced (50%).

- Description of Validation Dataset:

  Imbalance for pneumonia cases was maintained in the validation set, in which the positive/negative ratio was 1:3.

5. Ground Truth

This NIH Chest X-ray Dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. To create these labels, the authors used Natural

Language Processing to text-mine disease classifications from the associated radiological reports. The labels are expected to be >90% accurate and suitable for weakly-supervised learning.

The benefits of this ground truth are the extensive database built based on real data extracted from a clinical flow environment, which includes all the potential difficulties to be found and overcome by the algorithm in its implementation. The limitations are the lack of accuracy in at least 10% of labels and the lack of confirmation by other means (expert consensus, tissue studies, or any other) of the diagnosis.

6. FDA Validation Plan

- Patient Population Description for FDA Validation Dataset:
  A balanced gender dataset of chest x-rays in AP view of healthy subjects and pneumonia patients with ages between 18-65 years reflective of the incidence of pneumonia in a clinical setting.

- Ground Truth Acquisition Methodology:
  The optimal silver standard approach for this case is to use several radiologists (3) to label the images.

- Algorithm Performance Standard:
  As indicated in the suggested paper, an F1 score higher than 0.4. The selected F1 score for the submitted model is 0.81