# FDA  Submission

Carl Franklin

Pneumonia Screening Algorithm

## Algorithm Description

### 1. General Information

This algorithm is intended for assisting the radiologist to detect pneumonia cases in men and women from the ages of 10-90 who have been administered a screening chest x-ray study on a digital projection radiography imaging device.

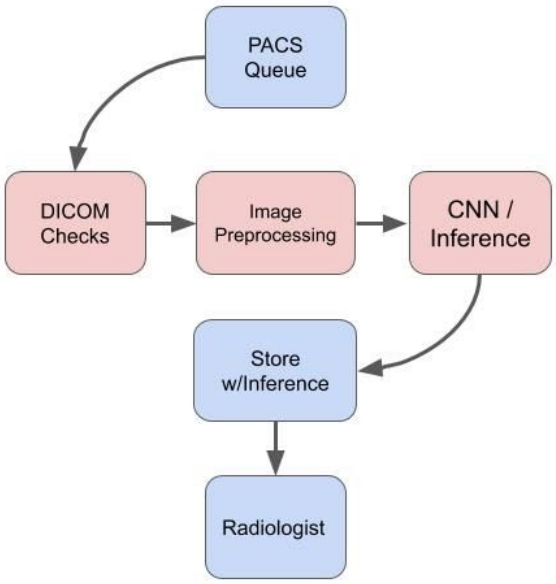
The algorithm can be used to identify patients requiring expedited analysis by a radiologist for diagnosis of pneumonia. Only chest x-rays are used by the algorithm to identify pneumonia with posterioranterior (PA)  and anteroposterior (AP) view positions.

The algorithm should only be used for screening on pneumonia and not other pulmonary conditions. The algorithm should only be used for screening to assist a radiologist and not for diagnosis.

The algorithm allows for assisting to prioritize the radiologist work queue to raise the priority of patients identified as having pneumonia. This will lead to faster identification of at risk patient receiving timely care.

## 2. Algorithm Design and Function

Flow chart showing original image workflow and additions for the algorithm design.




DICOM checking steps look to ensure that the DICOM file retrieved from the queue is appropriate to process with the algorithm. The DICOM checking verifies the following elements:

- Ensure the Modality is Digital Radiography (‘DX’)
- Ensure the BodyPartExamined is ‘CHEST’
- Verify PatientPosition is either posterioranterior (‘PA’)  and anteroposterior (‘AP’)

Image Preprocessing is performed to ensure the image is prepared for inferencing. The image is resized to a 224x224 image with a single monochrome color channel. The image is further scaled to have image pixel values scaledto between 0.0 and 1.0 from their original range of 0 to 255.

The CNN is configured with stored layers and layer weights associated with the device. The CNN is a modified version of the VGG-16 architecture initially loaded with ImageNet weights. There are several layers added to the end of the architecture and described in the training details below. The modified image is provided and a result of “No Pneumonia Finding’ or ‘Pneumonia Finding’. The algorithm provides this for screening purposes and prioritization of the Radiologists processing queue.

The algorithm finding is stored with the DICOM metadata back into the PACS system. The device may be used with this information to reprioritize the queue and allow the radiologist to optimize time to analysis for potential pneumonia patients.

## 3. Algorithm Training

Augmentation of the training data was performed to provide multiple image positions for the algorithm to learn. All images were scaled from 0 to 255 to 0.0 to 1.0 range. The following were the image augmentation parameters provided to allow the algorithm to see various images adjusted:

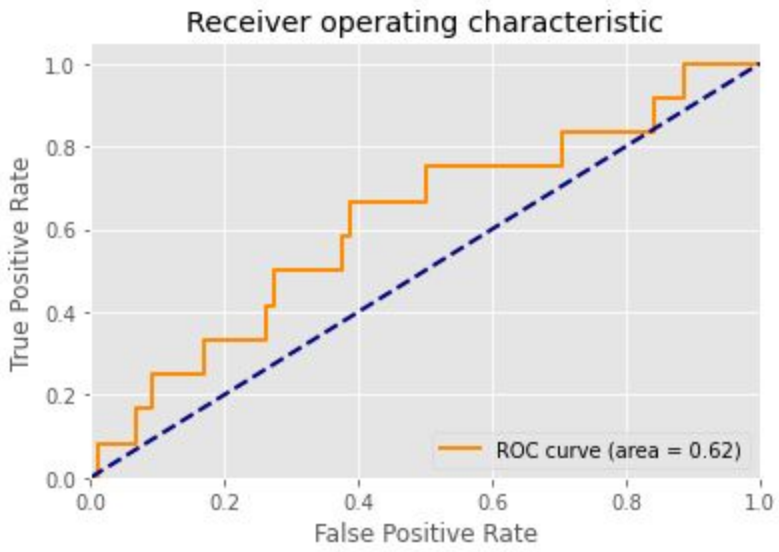
- Horizontal flip
- Shift height by up to 10%
- Shift width by up to 10%
- Rotate image up to 20 degrees
- Shear the image up to 10%
- Zoom the image up to 10%

Batch size was set at 100 to ensure a good mix in the validation batch that included pneumonia and non-pneumonia cases. The optimizer learning rate was set at 1e-3. The loss is set as ‘binary_crossentropy’ and the metric was set to ‘binary_accuracy’.

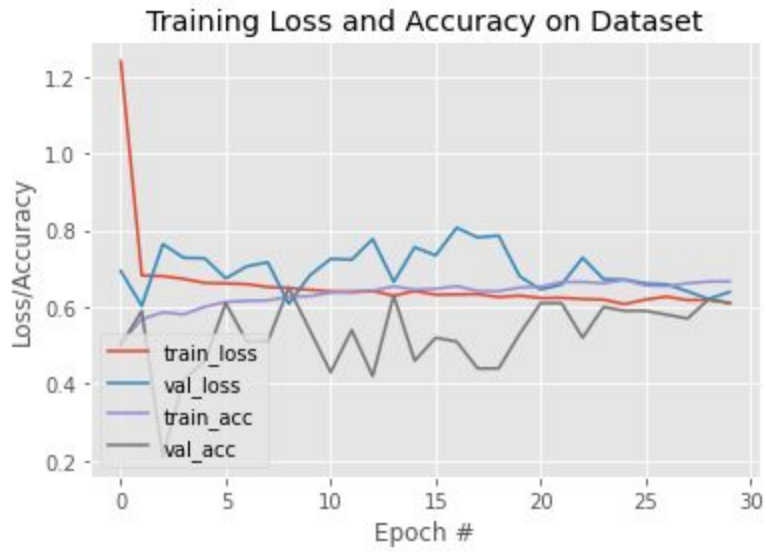
The CNN is a modified VGG-16 with the following modifications:

- The first 16 layers are frozen.
- Additional layers are added to the end for feature recognition
    - Flatten
    - Dropout with .03
    - Dense 1024 with relu activation
    - Dropout with .03
    - Dense 512 with relu activation
    - Dropout with .03
    - Dense 256 with relu activation
    - Dense 1 with sigmoid activation

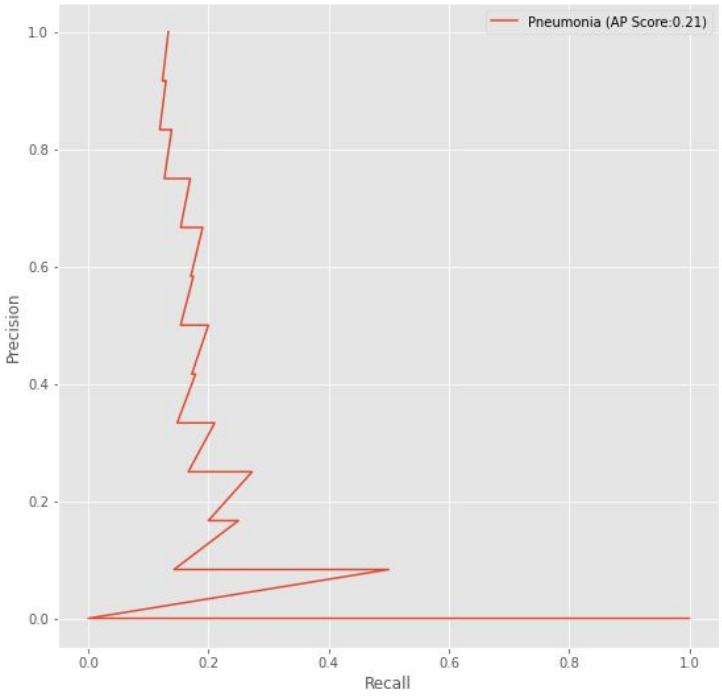
Information about training process



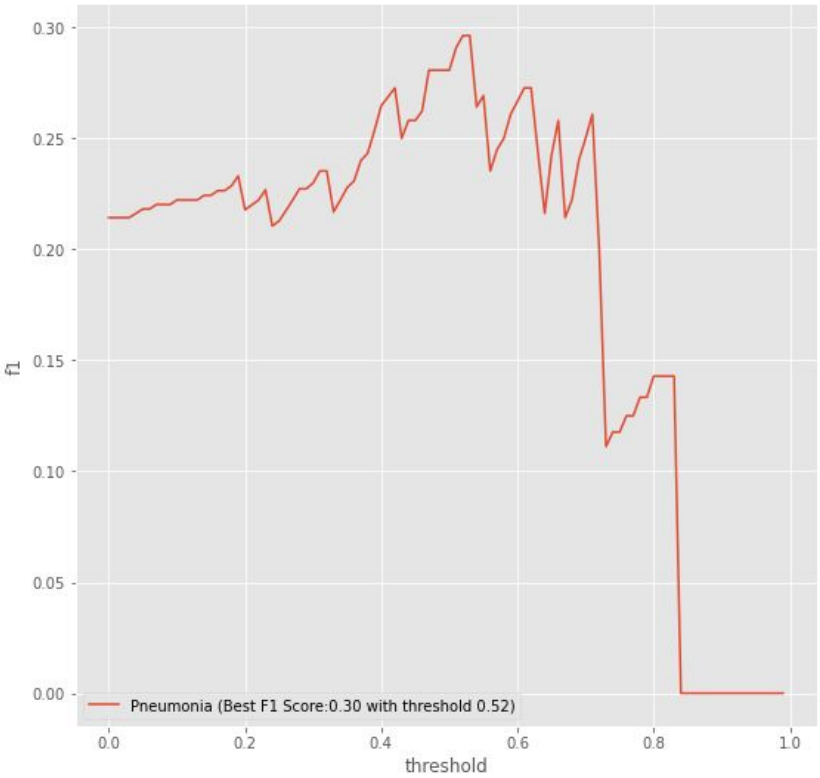

AUC is .62.




Steady training loss along with increasing training accuracy. Model does not appear to be overfitted to training data.




Precision-Recall curve shows precision/recall tightly coupled to a fairly narrow range of values with an AP of .21.

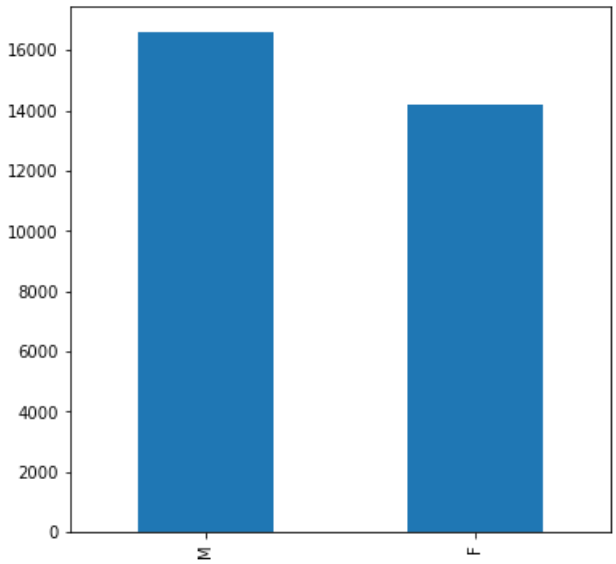


Best F1 scores are achieved between .4 and .7 which is a fairly wide range with a maximized peak at .52.

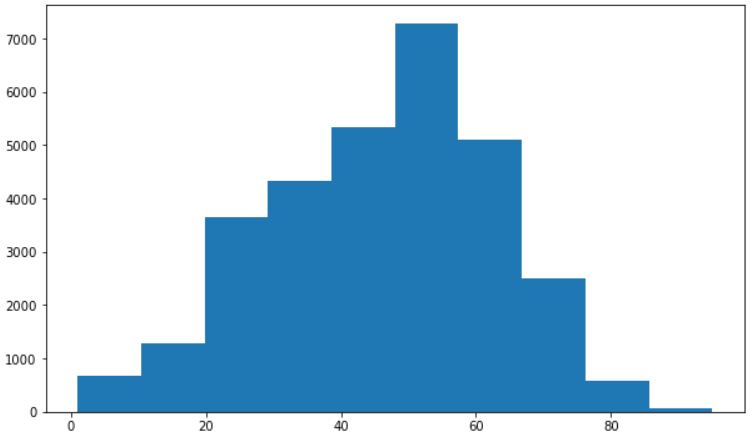
When evaluating for screening, it is more important to minimize false negatives and have a good balanced F1 score. With a high recall, nearly all images are labeled resulting in a high ratio of false positives. A better balance is the peak F1 score achieved at .52 threshold.

## 4. Databases

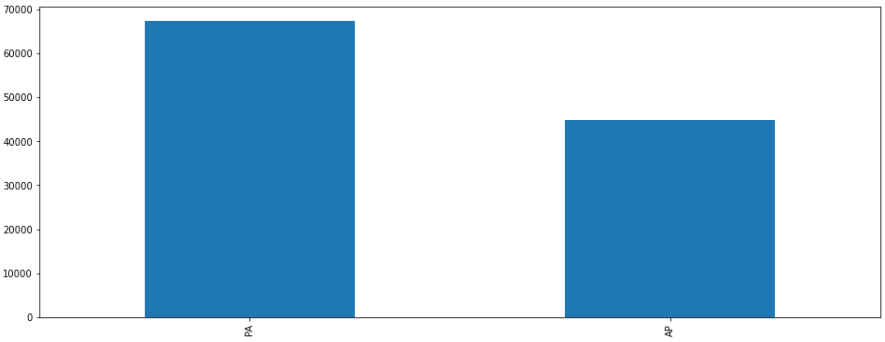
The database of images is the NIH Chest X-ray Dataset which contains 112,000 chest x-rays with disease labels acquired from 30,000 patients. The data was reviewed for relevant demographics to ensure proper alignment. Information on age, gender, view position, and occurrence of findings provided the guidance to intended use of the algorithm.



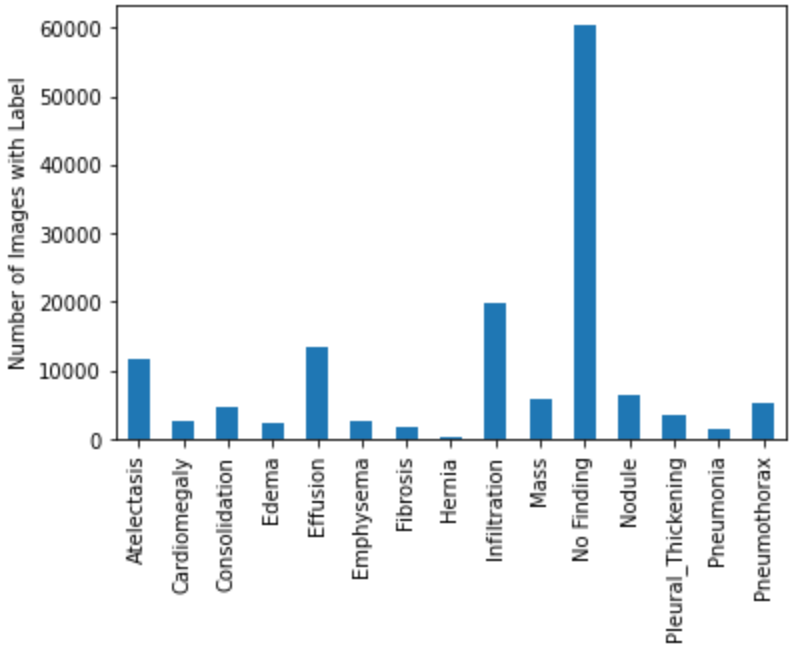

Analysis of Gender



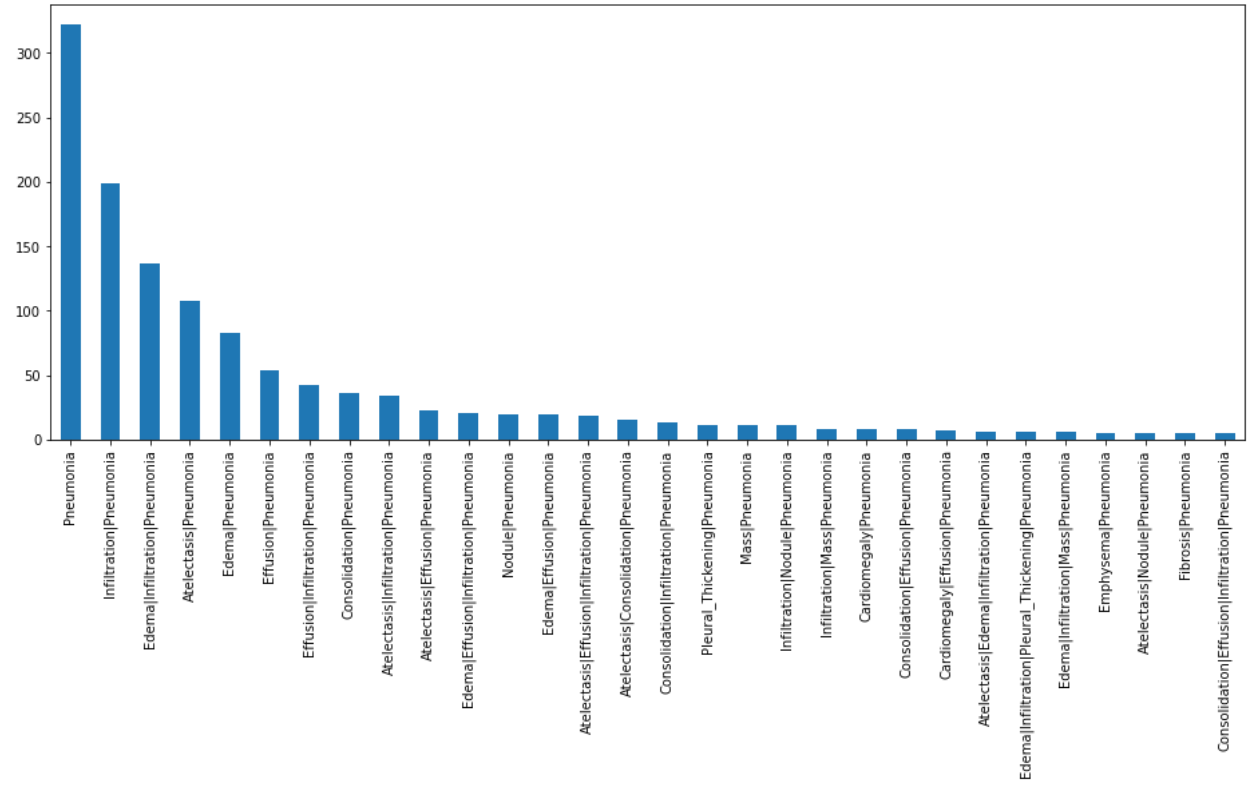

Age (adjusted to remove invalid records)




View Position of the x-ray



Findings in the dataset




Comorbid conditions with Pneumonia where the first column is only pneumonia findings.

The training dataset is initially taken as 80% of the NIH dataset then adjusted to ensure that the ratio of pneumonia to non-pneumonia is set to 50/50 to provide for training.

The validation dataset is initially taken as 20% of the NIH dataset. Then to adjust for accurately identifying false negative cases, the dataset is adjusted to 20/80 pneumonia/non-pneumonia. This adjustment ensures that labeling everything as non-pneumonia does not create a good precision/recall outcome.

## 5. Ground Truth

The radiology reports associated with the DICOM images are not publicly shared. Natural Language Processing (NLP) was used on the reports to extract labeling. There may be some errors but the accuracy of labels is estimated at > 90%. These label extractions are from radiologist reports and are the gold standard for the labeling of the images.

## 6. FDA Validation Plan

The patient population for the validation plan will include patients with the following information:

- Age between 10 and 90.
- Gender is male or female
- Image modality is digital radiography (‘DX’)
- Body part image is ‘CHEST’
- View position of the image is posterioranterior (‘PA’)  and anteroposterior (‘AP’)
- Pneumonia presence should be at least 20% of the validation data set

Because of the difficulty of identifying pneumonia, a silver standard leveraging the consensus of three radiologists would improve ground truth definition. The three radiologists assessments would be equally weighted to determine the proper finding of pneumonia or non-pneumonia.

The evaluation performance standard is to evaluate 3 (or more) radiologists against a subset of the validation data where a finding has been verified. Evaluate the F1 score (precision and recall) of each radiologist and combined to set the evaluation standard. The algorithm thereby would need to be within statistically insignificant F1 score (or higher) to be evaluated as successfully performing.