FDA Submission

Your Name:
Omar Laham Company

Name of your Device:
AI Algorithm for Pneumonia Detection in Chest-XRays

Algorithm Description

1. General Information

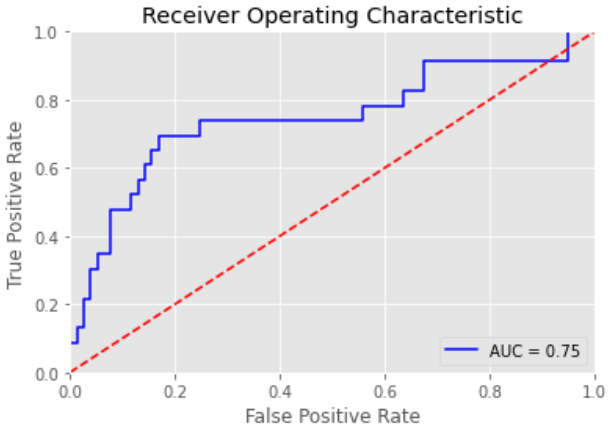
Intended Use Statement:
This algorithm is intended for use by radiologists on men and women from the age of 20-80 who have been administered a screening XRay study on a digital X-Ray machine to help them save time reading scans by prioritizing the X-Ray images according to the probability of including signs of Pneumonia.

Indications for Use:
This algorithms can be used on chest X-rays taken by digital X-Ray machines when the patient is in an 'AP' or a 'PA' position. It can be used for males and females in age of 20-80 years.

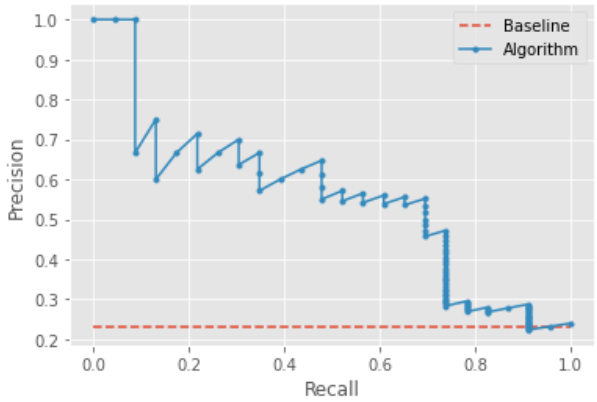
Device Limitations:
The algorithm doesn't provide extremely high performance and it can miss-predict when the image includes findings that look like findings in Pneumonia. However, this is something that even a radiologist find difficulty dealing with since detecting pneumonia from X-Ray is not easy even for a radiologist.

Clinical Impact of Performance:

roc_curve:



pr_curve:

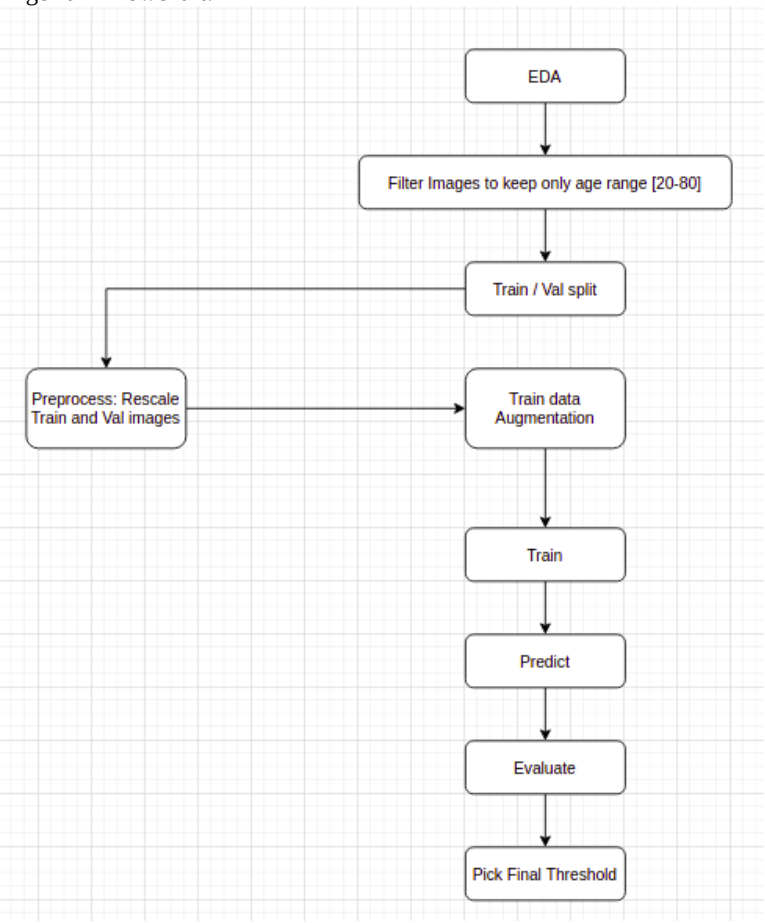


performance_statistics:

	Threshold	Sensitivity	Spec.	Precision	Recall	F1_score	Accuracy
0	0.0	1.000000	0.000000	0.230000	1.000000	0.373984	0.23
1	0.1	1.000000	0.012987	0.232323	1.000000	0.377049	0.24
2	0.2	0.913043	0.155844	0.244186	0.913043	0.385321	0.33
3	0.3	0.826087	0.337662	0.271429	0.826087	0.408602	0.45
4	0.4	0.739130	0.597403	0.354167	0.739130	0.478873	0.63
5	0.5	0.695652	0.805195	0.516129	0.695652	0.592593	0.78
6	0.6	0.565217	0.870130	0.565217	0.565217	0.565217	0.80
7	0.7	0.304348	0.961039	0.700000	0.304348	0.424242	0.81
8	0.8	0.086957	0.987013	0.666667	0.086957	0.153846	0.78
9	0.9	0.000000	1.000000	NaN	0.000000	NaN	0.77

2. Algorithm Design and Function

Algorithm flowchart:



DICOM Checking Steps:
1- Modality: should be “DX”, otherwise it prints an error and doesn’t predict.
2- Sex: should be either ‘M’ or ‘F’, otherwise it prints a warning
3- Patient Age: should be between 20 and 80 (included), otherwise it prints a warning.
4- Body Part Examined: must be “CHEST”, otherwise it prints an error and doesn’t predict.
5- Patient Position: must be either ‘AP’, or ‘PA’, otherwise it prints an error and doesn’t predict.

Preprocessing Steps:

- Rescale image intensity by 1/255
- Normalization: we use a mean of 0 and a standard deviation of 1. This means that normalization is not actually used since this will have zero effect on the image.

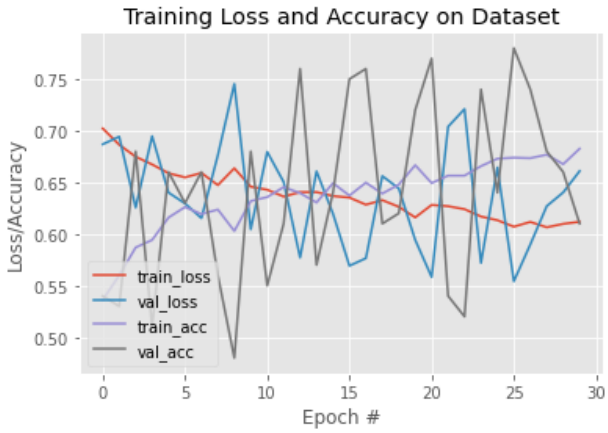
CNN Architecture:
VGG16 –> Flatten -> Dense(512) –> Dropout(0.3) -> Dense(256) –> Dense(1)

3. Algorithm Training

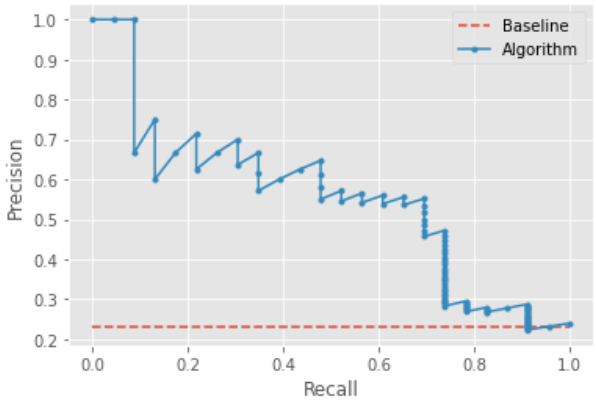
Parameters:

- Types of augmentation used during training
  Horizontal flip, height_shift in range of 0.1, width_shift in range of 0.1, ration in range of 20 degrees, shearing in range of 0.1, zooming in range of 0.1.
- Batch size: 16
- Optimizer learning rate: 1e-5
- Layers of pre-existing architecture that were frozen: First 17 layers
- Layers of pre-existing architecture that were fine-tuned: None
- Layers added to pre-existing architecture:
  - Flatten layer
  - Dense Layer (Relu, 512)
  - Dropout Layer (0.3)
  - Dense Layer (Relu, 256)
  - Dense Layer (Sigmoid, 1)

Algorithm training performance visualization:



P-R curve:

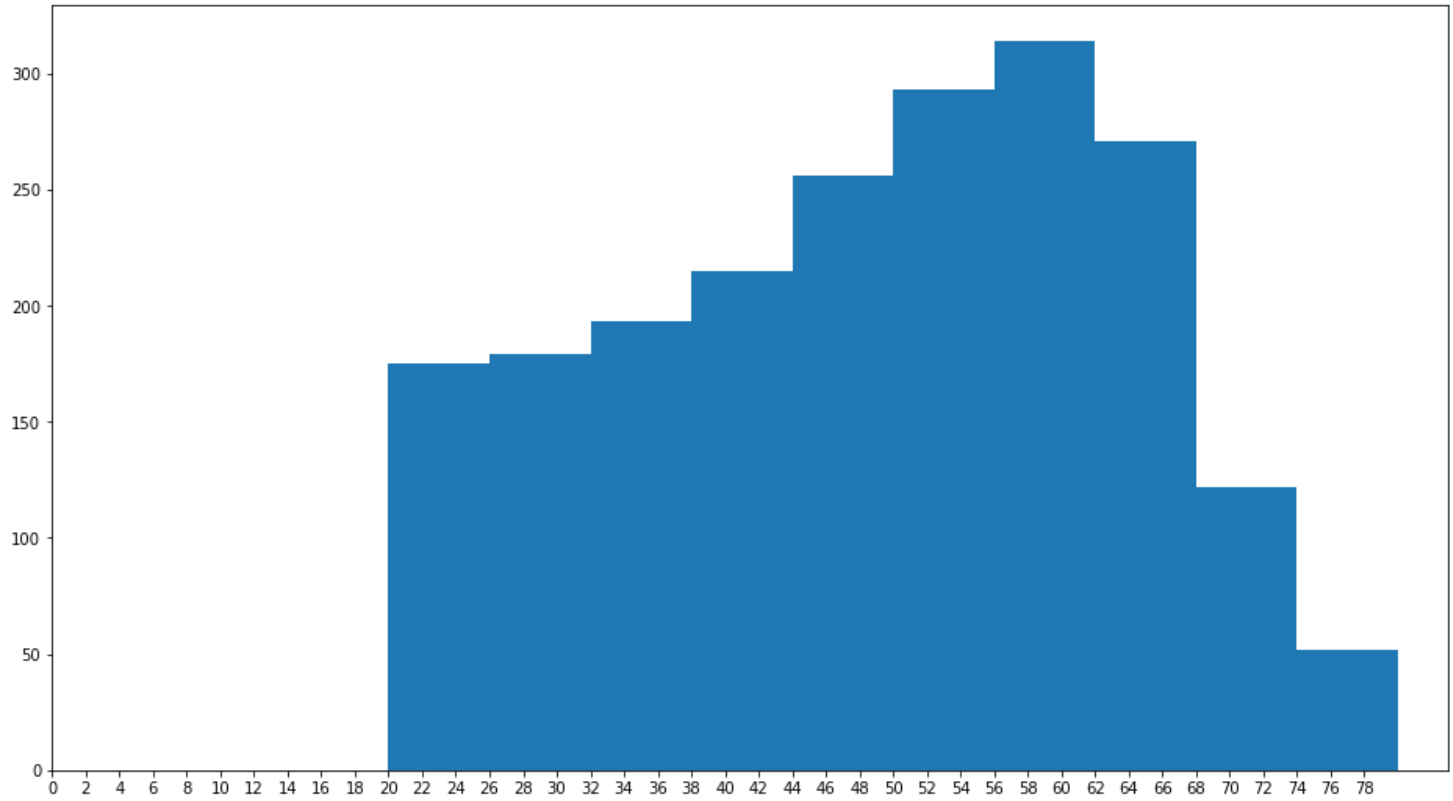


Final Threshold and Explanation:
The threshold was picked according to the the best F1 score in balance with accuracy.
Final probability threshold picked to determine if an image contains pneumonia signs is: 0.6.

4. Databases

Description of Training Dataset:
Number of images:2070
N Positive Pneumonia:1035
N Negative Pneumonia:1035

Age Histogram:

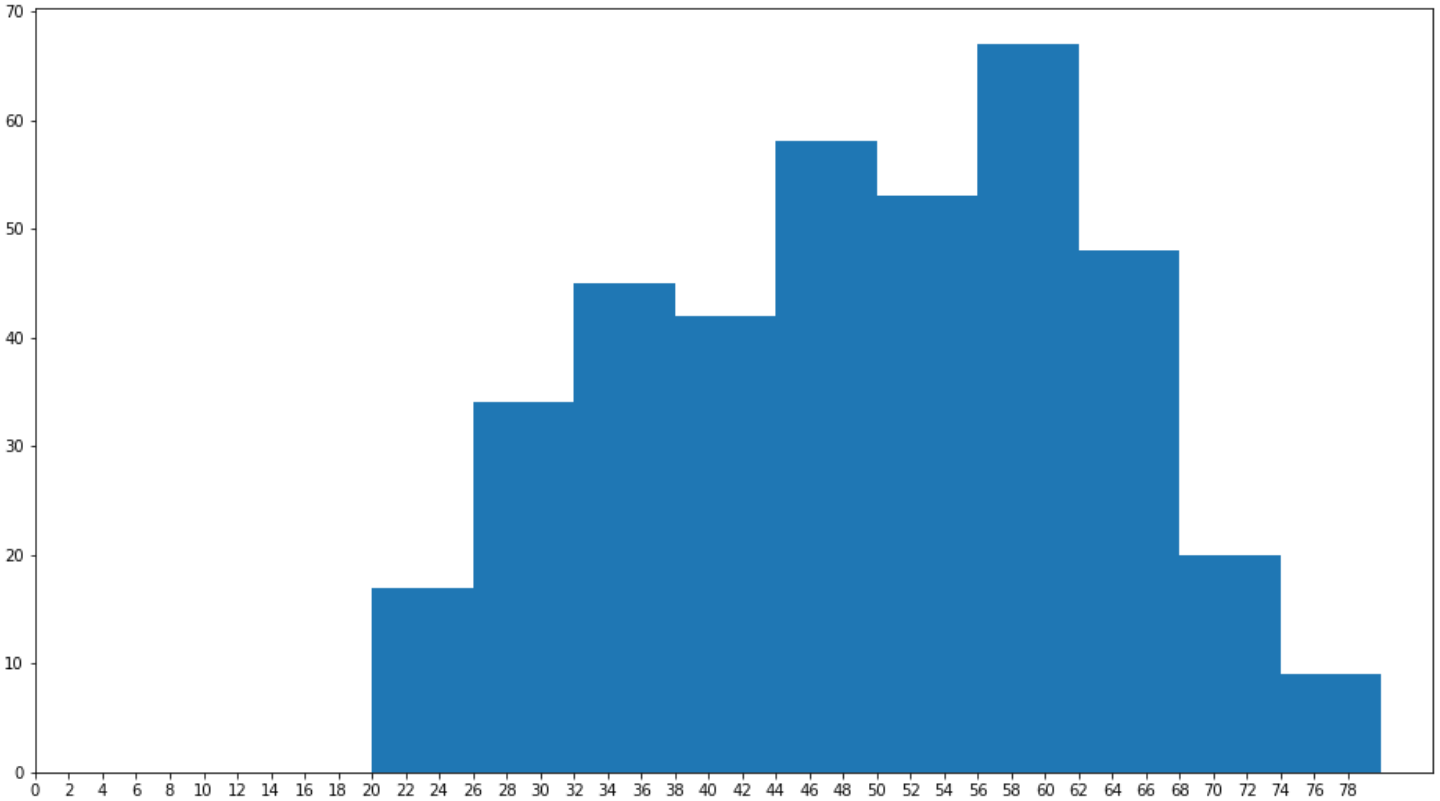


Sex Barplot:

Position Barplot:

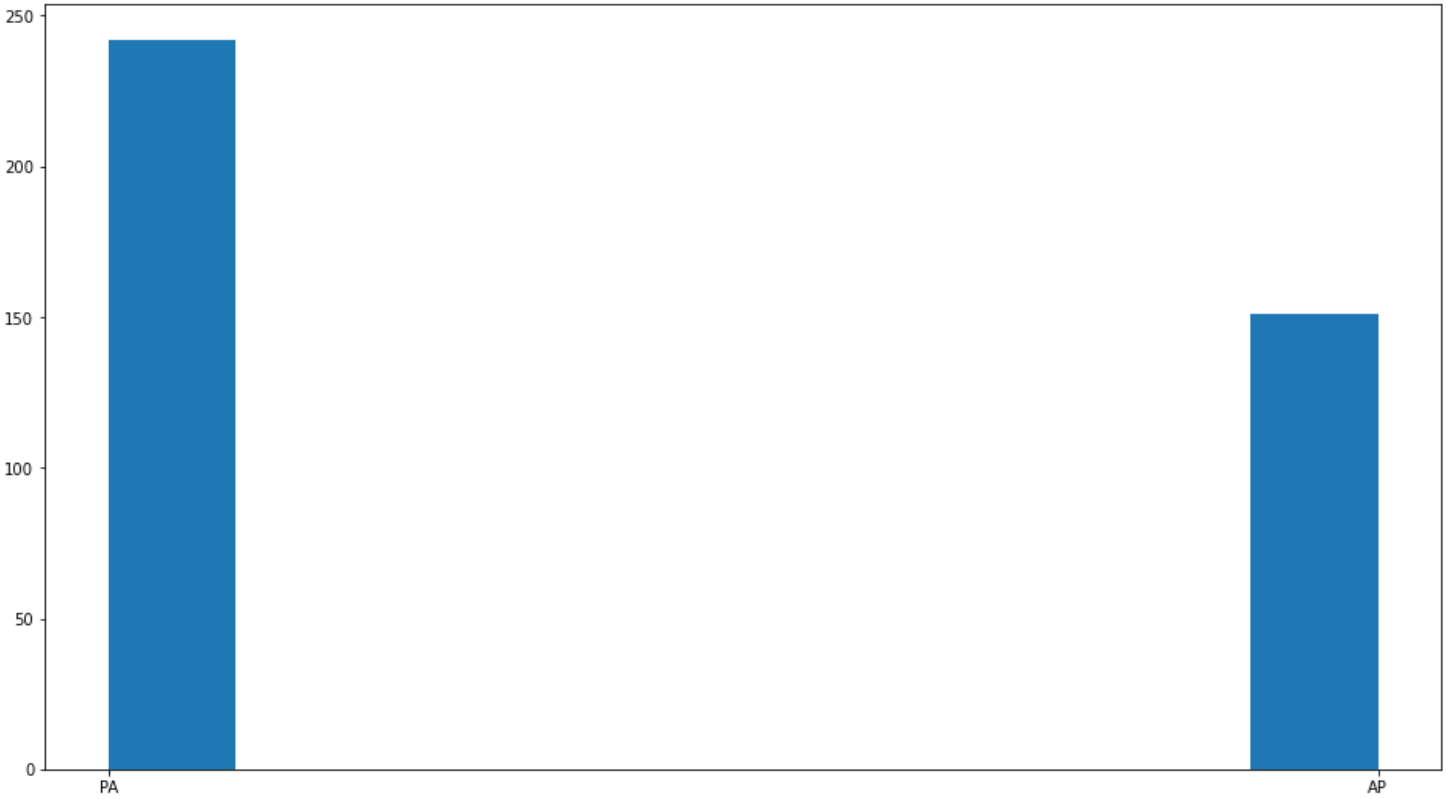
Description of Validation Dataset:
Number of images:399
N Positive Pneumonia:4 (1%)
N Negative Pneumonia:399

Age Histogram:



Sex Barplot:

Position Barplot:



5. Ground Truth

Since the algorithm is only ment to help the radiologist and not replace him / her, then a single radiologist's label extracted from the Radiology reports accompanying each image is sufficent as a ground truth, especially that the algorithm is performing better than the average F1 score of 4 four practicing academic radiologists as will be listed under "Algorithm Performance Standard" section.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

A validation set that was made up of DX studies for men and women between the ages of 20 and 80.

To validate the algorithm, we want to make sure that my validation data set has a similar population distribution as the training dataset in terms of sex, age and view position.

Ground Truth Acquisition Methodology:
Since the algorithm is only ment the radiologist and not replace him / her, then a single radiologist's labels would probably suffice as a ground truth, especially that the algorithm is performing better than the average F1 score of 4 four practicing academic radiologists as will be listed under next section.

Algorithm Performance Standard:
For your validation plan, you need evidence to support your reasoning. As a result, you need a performance standard. This step usually involves a lot of literature searching.

This article https://arxiv.org/pdf/1711.05225.pdf contains the F1 score of "Radiologist-Level Pneumonia Detection on Chest X-Rays" for 4 four practicing academic radiologists. the mean F1 score is 0.387.
This algorithm has an F1 score of 0.424242 which is better than the average F1 score of 4 four
practicing academic radiologists.

The algorithm is taking ~ 20 milliseconds to predict a probability of how sure it is about the presence of pneumonia signs in the X-Ray which makes it extremely helpful to priorotize the queue of X-Ray images that a radiologist should read.