FDA Submission

Your Name:

IRFAN MANSURI

Name of your Device:

X-Ray Pneumonia Classifier

Algorithm Description

1. General Information

Intended Use Statement:
For assisting a radiologist in detection of detection of pneumonia in X-Ray images based on CNN architecture.

Indications for Use:

For assisting a radiologist in detection of pneumonia in x-ray images.

Targeted Patient Population:

- Both men and women
- Age: 1 to 100

X-Ray Image Properties

- Body part: Chest
- Position: AP (Anterior/Posterior) or PA (Posterior/Anterior)
- Modality: DX (Digital Radiography)

There could be certain comorbities with pneumonia like Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pleural thickening, Cardiomegaly, Nodule, Mass, Hernia.

Device Limitations:

CPU with at least 8GB RAM

Clinical Impact of Performance:

The algorithm is trained to get a high F1 score and recall value. So the false positive rate maybe high, which should be considered when applied.

- False positives classify a patient with no pneumonia as positive to get unnecessary treatment;
- False negatives classify a patient with pneumonia as negative to miss required treatment, which is worse

2. Algorithm Design and Function

Algorithm Flowchart:

Exploratory Data Analysis -> Data Pre-processing -> Feature Engineering -> Model Training -> Prediction and Evaluation

DICOM Checking Steps:

The Algorithm performs the following checks on the DICOM image:

- Check Patient Age is between 1 to 100.
- Check Examined Body Part is 'CHEST'.
- Check Patient Position i.e. either AP (Anterior/Posterior) or PA (Posterior/Anterior).
- Check Modality is 'DX' (Digital Radiography)

Pre-processing Steps:

The Algorithm performs the following pre-processing steps on an image data:

- Convert RGB to Grayscale (iff needed).
- Re-sizes the image
- Normalizes the intensity

CNN Architecture:

Below is the CNN architecture graph.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
model (Model)                (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dropout (Dropout)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 1024)              25691136
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_1 (Dense)              (None, 512)               524800
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_2 (Dense)              (None, 256)               131328
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 257
=================================================================
Total params: 41,062,209
Trainable params: 28,707,329
Non-trainable params: 12,354,880
_________________________________________________________________
```

2. Algorithm Training

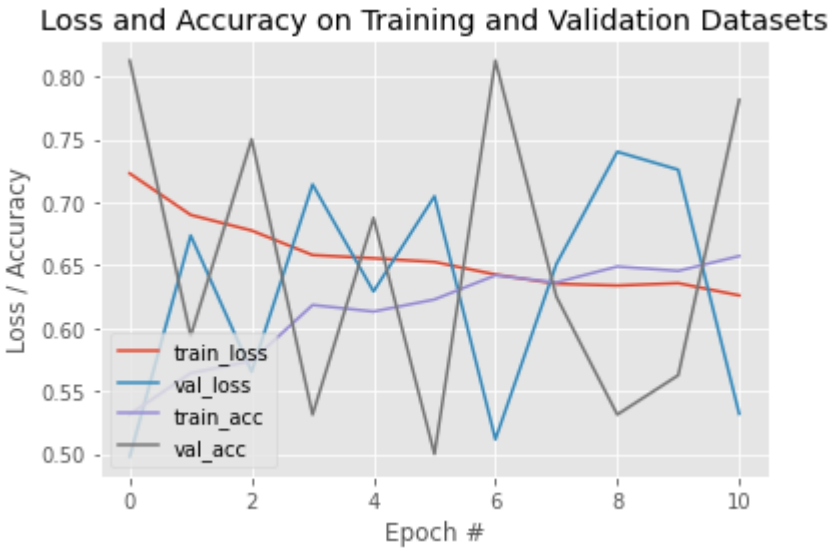
Parameters:

- Types of augmentation used during training
  Horizontal Flip:
  Height Shift: 0.1
  Width Shift: 0.1
  Rotation Angle Range: 0 to 20 degrees.
  Shear: 0.1
  Zoom: 0.1

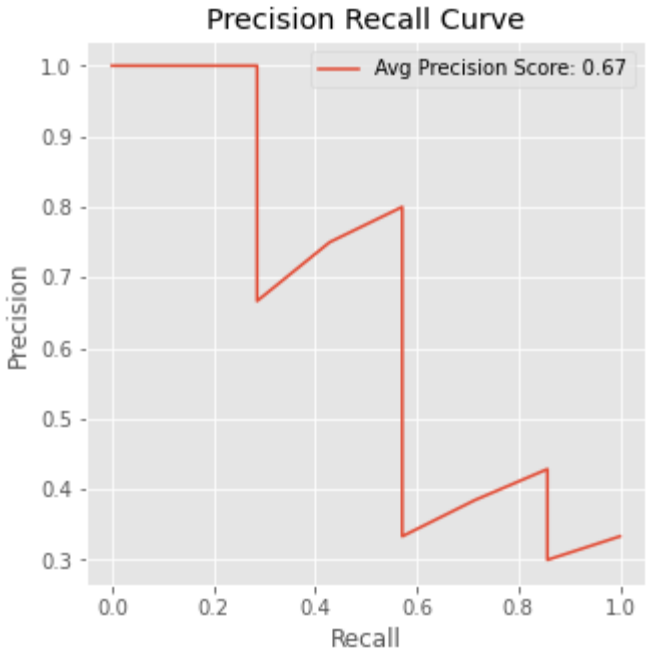
- Batch Size: 32
- Optimizer Learning Rate: 1e-4 (Adam Optimizer)
- Layers of pre-existing architecture that were frozen
  First 17 layers of pre-existing architecture are frozen
- Layers of pre-existing architecture that were fine-tuned
  The last layer of pre-existing architecture is fine-tuned
- Layers added to pre-existing architecture
  Flatten, Dropout, and Dense layers are added to pre-existing architecture, as shown in CNN Architecture above

Flatten, Dropout, and Dense layers are added to pre-existing architecture, as shown in CNN Architecture above

Algorithm Training Performance Visualization:

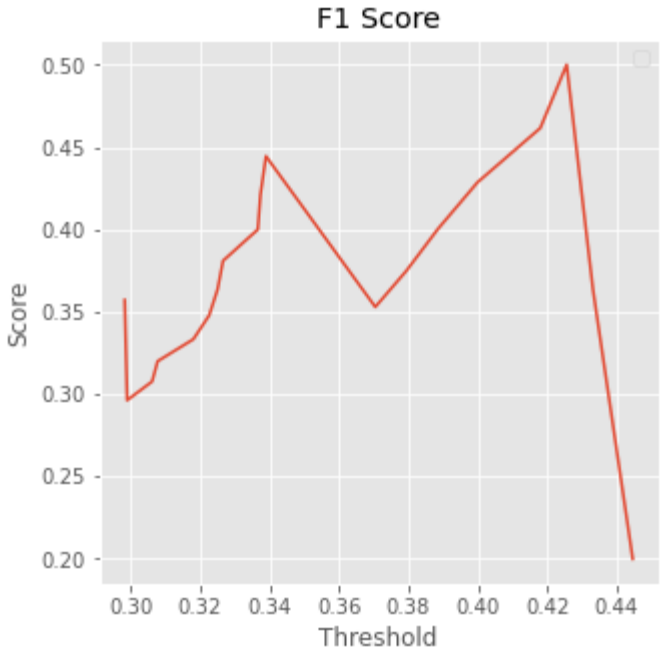


P-R Curve:



Precision is TP / (TP + FP), the percentage of true positives over all positive results, the higher precision, the more confident about positive results; recall is TN / (TN + FP), the percentage of true negatives over all patients with no disease, the higher recall, the more confident about negative results. Consider the trade-off between precision and recall, F1 score = 2 x precision x recall / (precision + recall) is selected as criterion.

Final Threshold and Explanation:

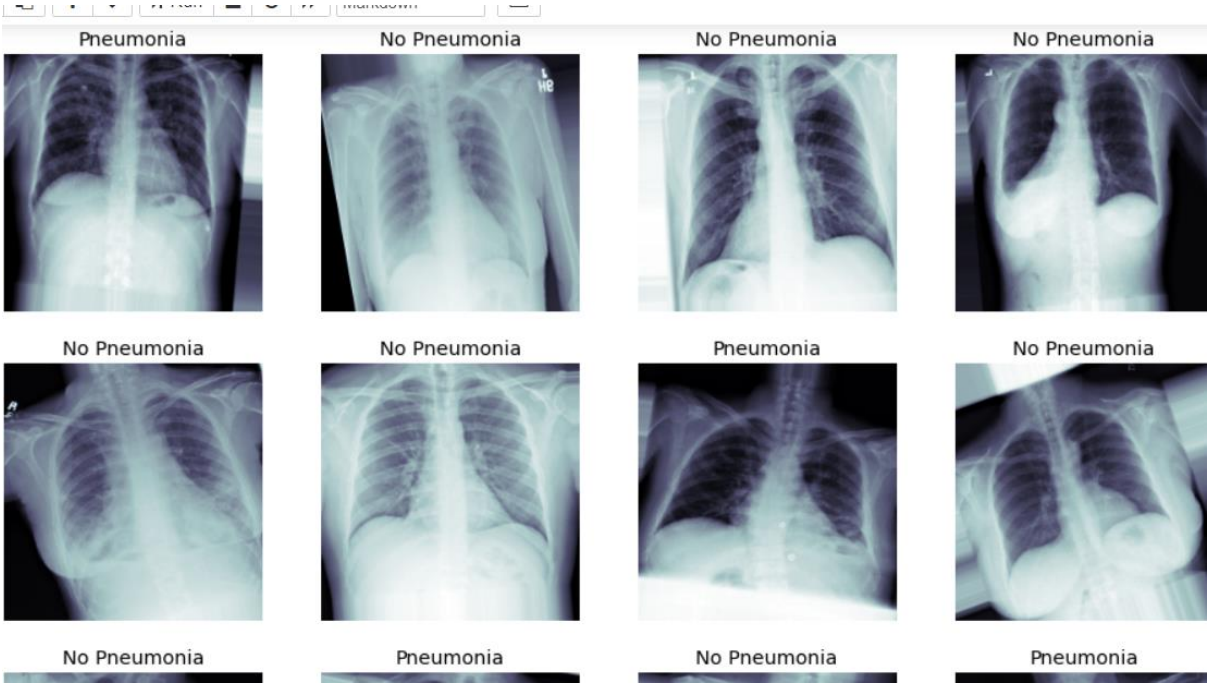


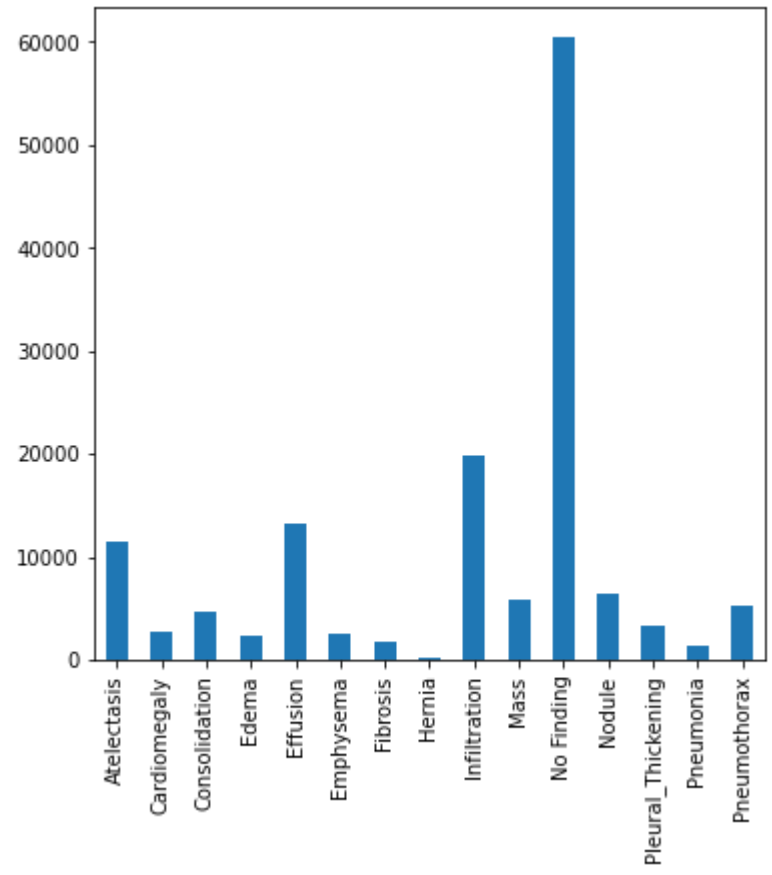
Final threshold is set as 0.4255593 to get higher F1 score 0.5

4. Databases

For the below, include visualizations as they are useful and relevant

There are 112,120 X-ray images with disease labels from 30,805 unique patients in this dataset. The disease labels were created using Natural Language Processing (NLP) to mine the associated radiological reports. The labels include 14 common thoracic pathologies:





From the disease distribution, Infiltration, Effusion, and Atelectasis are top 3 except No Finding.

Description of Training Dataset:

The percentage of the presence of pneumonia in the original training dataset is 0.013, and this imbalance is adjusted according to the lesson in the course to make the new percentage become 0.5

Description of Validation Dataset:

The percentage of the presence of pneumonia in the original validation dataset is also 0.013, and this unbalance is adjusted following the lesson in the course to become 0.2.

5. Ground Truth

The disease labels were created using Natural Language Processing (NLP) to mine the associated radiological reports. The biggest limitation of this dataset is that image labels were NLP-extracted so there could be some erroneous labels but the NLP labelling accuracy is estimated to be > 90%.

- Atelectasis
- Consolidation
- Infiltration
- Pneumothorax
- Edema
- Emphysema
- Fibrosis
- Effusion
- Pneumonia
- Pleural thickening
- Cardiomegaly
- Nodule
- Mass
- Hernia

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

For FDA Validation Dataset, the sampling should be based on equal number of male and female patients, the age range between 1 and 100, and including both the presence and absence of any of 14 common thoracic pathologies.

Ground Truth Acquisition Methodology:

To acquire Ground Truth, make sure Modality == ‘DX’ and PatientPosition in [‘AP’, ‘PA’] and BodyPartExamined == ‘CHEST’, and prefer biopsy data.

Algorithm Performance Standard:

Metrics chosen to monitor for Performance could be F1 score, Precision, or Recall, and F1 score is a better selection. According to the paper "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning" (https://arxiv.org/pdf/1711.05225.pdf), CheXNet's F1 score is 0.435. After applying the current algorithm to FDA Validation Dataset, this model would output a F1 score, then the statistical significance of the new score compared to the CheXNet's score should be checked. To support that this algorithm is a better one, the new score is not necessary to be equal or higher, it could be lower but inside a tolerance window, in this case, the calculated p-value should be compared to the confidence level or 95% confident interval should be examined, all conclusion should be based on statistical inference!