FDA Submission

Your Name: Jamiu Ekundayo

Name of your Device: cheXray

Algorithm Description

1. General Information

Intended Use Statement: This algorithm is intended for use in assisting a radiologist with pneumonia identification from chest x-rays.

Indications for Use: It is indicated for use in patients (male and female) within the age bracket 1-95 years with chest x-rays taken in the AP and PA view positions on a ER setting.

Device Limitations: The presence of edema, atelectasis and effusion may reduce the performance of the algorithm in accurately predicting the presence of pneumonia in a chest x-ray. Conversely, the presence of infiltration and consolidation in the image may lead to improved performance of the algorithm at accurately predicting the presence of pneumonia from a chest x-ray.

Clinical Impact of Performance: This algorithm's performance shows that it will be useful for screening chest x-rays for pneumonia and may also be used for workflow prioritization.

2. Algorithm Design and Function
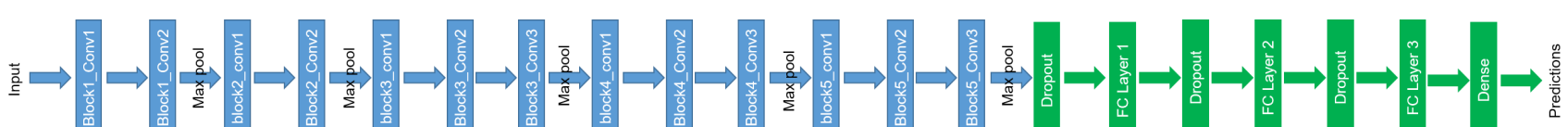
<< Insert Algorithm Flowchart >>
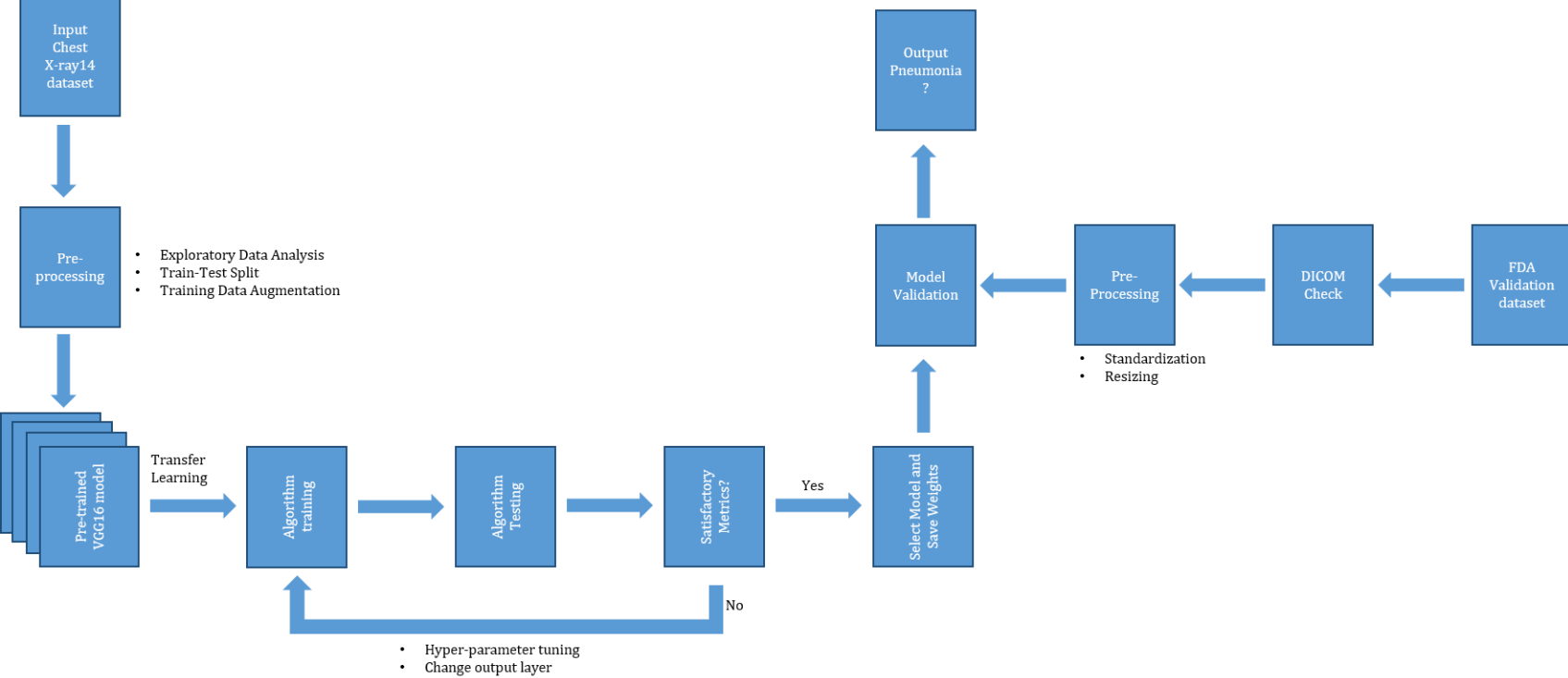
DICOM Checking Steps:

1. Checked the dicom file to ensure the body part examined was chest
2. Checked the dicom file to ensure the imaging modality was DX
3. Checked the dicom file to ensure the view positions were AP and PA

Preprocessing Steps:

1. Image standardization
2. Image resizing

**CNN Architecture:**

3. Algorithm Training

Parameters:

- Types of augmentation used during training - horizontal flip with 10% shift range in height, width, shear and zoom. Rotation range was set to 20 degrees.
- Batch size - 32 used for training; batch_size = length of validation dataset used for testing.
- Optimizer learning rate - 0.0001
- Layers of pre-existing architecture that were frozen - All output layers except output layer 10 (block3_conv3)
- Layers of pre-existing architecture that were fine-tuned - output layer 10 (block3_conv3)
- Layers added to pre-existing architecture - dropouts and three fully connected layers 1024, 512 and 256

<< Insert algorithm training performance visualization >>
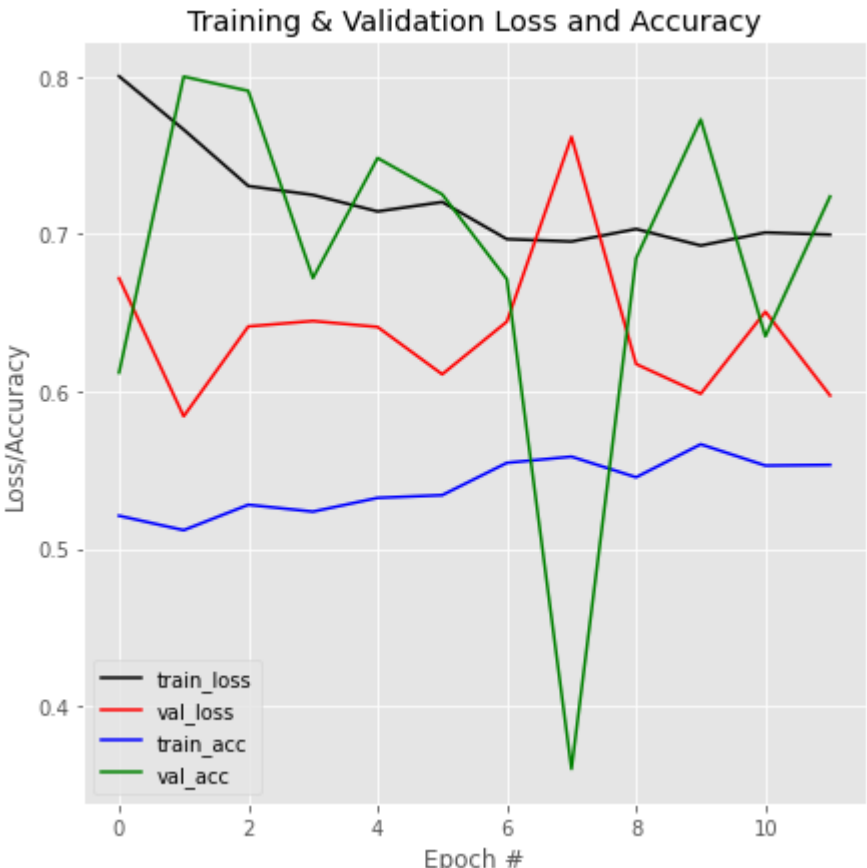
<< Insert P-R curve >>
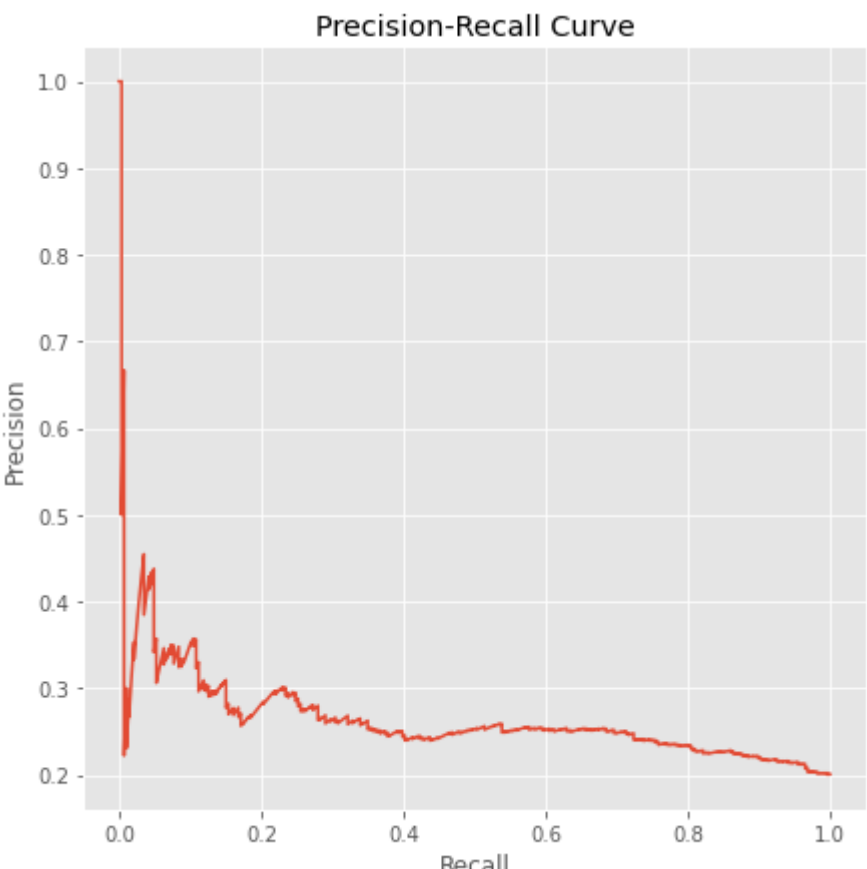
Final Threshold and Explanation: Final threshold = 0.387. This was the threshold that maximized F1_score (@ F1 score = 0.371). It corresponded to a 25% precision and 71% recall.

4. Databases

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

Description of Training Dataset: The training dataset is balanced for pneumonia and contains 2290 images sampled from 112,120 chest X-ray images with 15 (unique) disease labels from 30,805 unique patients.

Description of Validation Dataset: The validation dataset is an imbalanced dataset containing 20% pneumonia cases and a total of 1430 images sampled from 112,120 chest X-ray images with 14 (unique) disease labels from 30,805 unique patients.

5. Ground Truth

The labels were generated through text-mining radiologists' reports using a Natural Language Processing (NLP) algorithm. The NLP algorithm was reported to achieve over 90% accuracy. Moreover, this approach is fast and cost effective but it has inherent issues which might affect the accuracy of the algorithm and limit its clinical applications. For example,

```
1. the labels might not reflect the medical histories and/or blood test results of the patient
s
2. possibility of image-label mismatch.
```

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

My ideal dataset (from my clinical partner) for FDA validation would be chest x-ray images taken in the AP and PA view positions to check for pneumonia in both male and female patients within 1 - 95 years age bracket. The images may contain infiltration and consolidation (since these improve the algorithm performance at predicting pneumonia) but they must not contain edema, atelectasis and effusion since these reduce the algorithm's performances at accurately at predicting pneumonia from chest x-rays.

Ground Truth Acquisition Methodology:

The ground truth acquisition method would be a silver standard (visual ground truth) where the labels are acquired from expert consensus of three to four certified radiologists, with the votes weigthed by the years of experience of each radiologist. While it is known that this method may result large inter-observer variability in the labels (1), it is cheaper and faster to acquire than biopsy-based labelling. Moreover, the use of expert consensus approach is preferred to automated NLP labelling which is known to result in label noise (2, 3, 4) or image-label mismatch.

Algorithm Performance Standard:

With AUC = 0.602 and maximum F1_score = 0.371, CheXray performs lower than CheXNet (1). However, CheXray's F1 score is higher those of two radiologists (radioplogist 2 & radiologist 3) in Rajpurtar, et al. (2017) (1).

References:

1. https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/
2. https://arxiv.org/pdf/1711.05225.pdf
3. Tang, Y., Tang, Y., Peng, Y. et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. npj Digit. Med. 3, 70 (2020). https://doi.org/10.1038/s41746-020-0273-z
4. https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/