FDA Submission

Your Name: Karthik Rao

Name of your Device: Model to detect Pneumonia from Chest X-Ray scans

Algorithm Description

1. General Information

Intended Use Statement:

This model is based on Convolutional Neural Network algorithm trained to detect the presence of Pneumonia from Chest X-Rays scans. It is intended to be used in a clinical setting as an assist to radiologists while diagnosing Chest X-Rays scans.

Indications for Use:

The model may be used for screening patients suspected of having Pneumonia. Patient can of either gender and aged between 16 and 90. Also, the X-Ray images fed to the model must be of 'PA' or 'AP' viewing position.

Device Limitations:

In order to perform optimally the model needs to be run on a computer with fairly capable CPU and GPU.

Also, presence of 'Infiltration' may cause the model to classify an X-Ray image as positive for Pneumonia, thereby affecting the model's performance.

Clinical Impact of Performance:

The model aims to assist radiologists with their diagnosis by detecting presence of Pneumonia, hence, potentially speeding up screening and reducing fatigue.

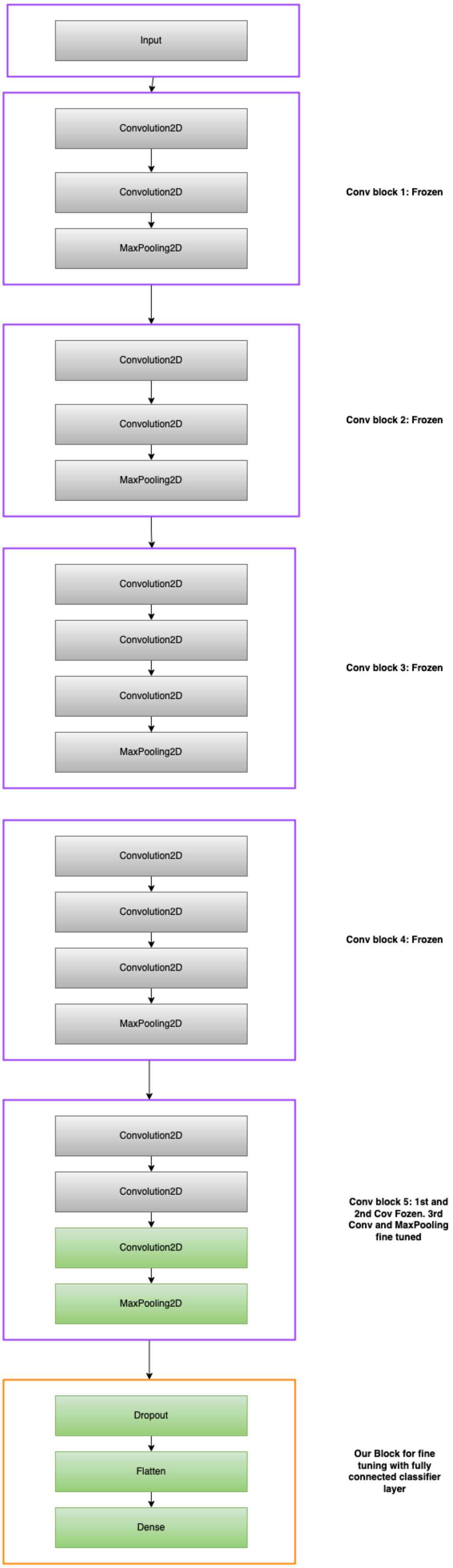
If the model predicts a false positive, the radiologist may advise the patient's sputum culture to be analysed for confirmation. Since this algorithm is optimised to have high recall, making it best suited for screening Pneumonia, a higher number of false positive prediction could be observed.

On the other hand, if the model predicts a false negative, it would adversly affect the patient. Hence, the radiologist should rely on his diagnosis, patient's medical history and could advise a sputnum culture.

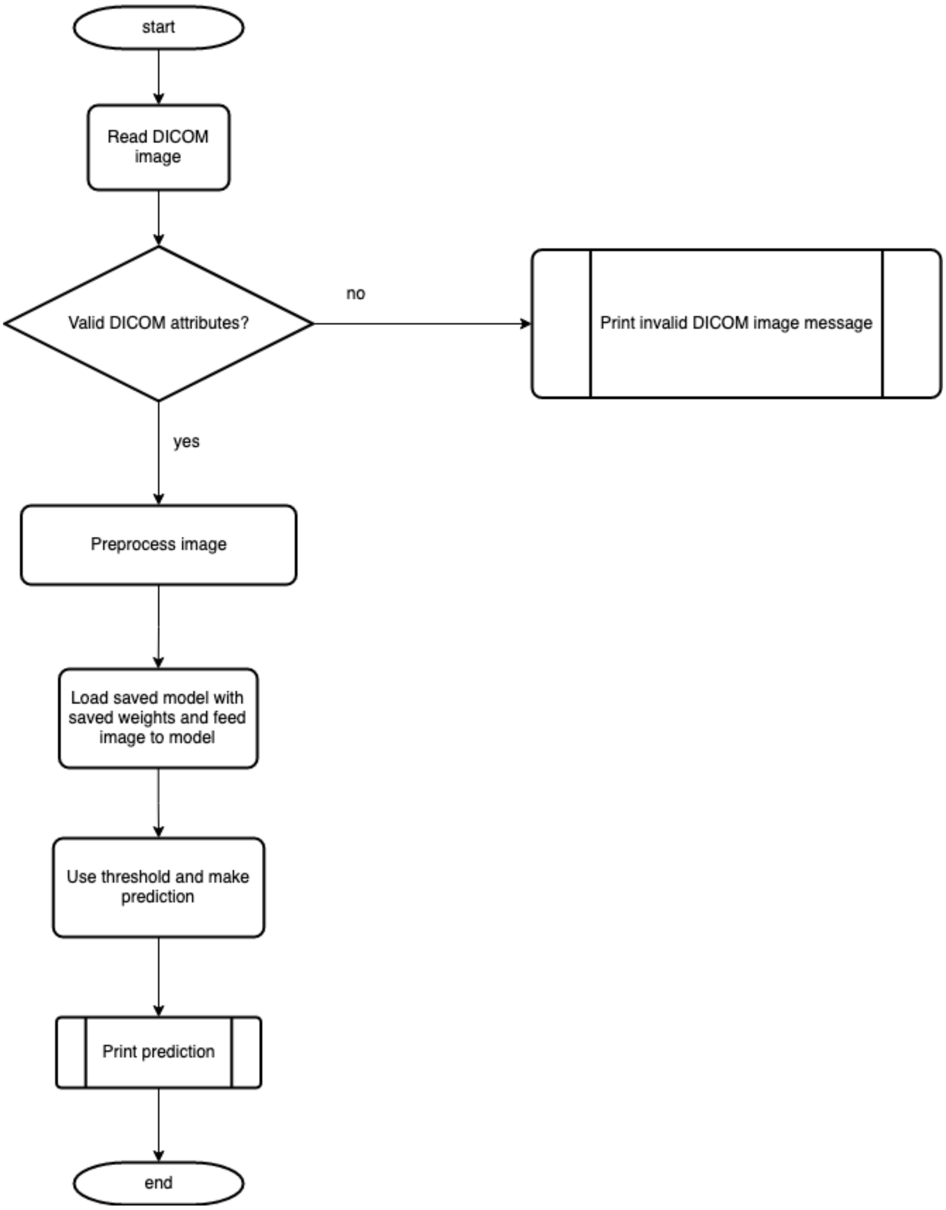
Therefore, a radiologist is always required to review the X-Ray images and validate the results.

2. Algorithm Design and Function

Model architecture:



Algorithm flowchart:



DICOM Checking Steps:

Following attributes are checked in DICOM Headers:

- Modality is 'DX'
- BodyPartExamined is 'CHEST'
- PatientPosition is either 'PA' or'AP'

If any of these three checks fail, then a message will be printed that DICOM images does not meet required criteria.

Preprocessing Steps:

Following are the sequence of steps:

- Perform checks are on DICOM header
- Extract DICOM pixel array
- Resize to 224 x 224
- Normalize the pixel array before feeding to model

CNN Architecture:

A sequential model was built that extends a pre-trained VGG16 model architecture with ImageNet weights. All the layers of VGG16 architecture upto layer 'block5_pool' are utilized. Except for the last convolution and last pooling layer all the layers are frozen, to prevent retraining their weights. To this architecture a flattening, dropout and dense layer are added. The final dense layer outputs a single class.

Layers from VGG16:

Layer (type) (Output Shape) Param

input_1 (InputLayer) (None, 224, 224, 3) 0

---

block1_conv1 (Conv2D) (None, 224, 224, 64) 1792

---

block1_conv2 (Conv2D) (None, 224, 224, 64) 36928

---

block1_pool (MaxPooling2D) (None, 112, 112, 64) 0

---

block2_conv1 (Conv2D) (None, 112, 112, 128) 73856

---

block2_conv2 (Conv2D) (None, 112, 112, 128) 147584

---

block2_pool (MaxPooling2D) (None, 56, 56, 128) 0

---

block3_conv1 (Conv2D) (None, 56, 56, 256) 295168

---

block3_conv2 (Conv2D) (None, 56, 56, 256) 590080

---

block3_conv3 (Conv2D) (None, 56, 56, 256) 590080

---

block3_pool (MaxPooling2D) (None, 28, 28, 256) 0

---

block4_conv1 (Conv2D) (None, 28, 28, 512) 1180160

---

block4_conv2 (Conv2D) (None, 28, 28, 512) 2359808

---

block4_conv3 (Conv2D) (None, 28, 28, 512) 2359808

---

block4_pool (MaxPooling2D) (None, 14, 14, 512) 0

---

block5_conv1 (Conv2D) (None, 14, 14, 512) 2359808

---

block5_conv2 (Conv2D) (None, 14, 14, 512) 2359808

---

block5_conv3 (Conv2D) (None, 14, 14, 512) 2359808

---

block5_pool (MaxPooling2D) (None, 7, 7, 512) 0

Layers added:

Layer (type) (Output Shape) Param

dropout_1 (Dropout) (None, 7, 7, 512) 0

---

flatten_1 (Flatten) (None, 25088) 0

---

dense_1 (Dense) (None, 1) 25089

Total params: 14,739,777

Trainable params: 2,384,897

Non-trainable params: 12,354,880

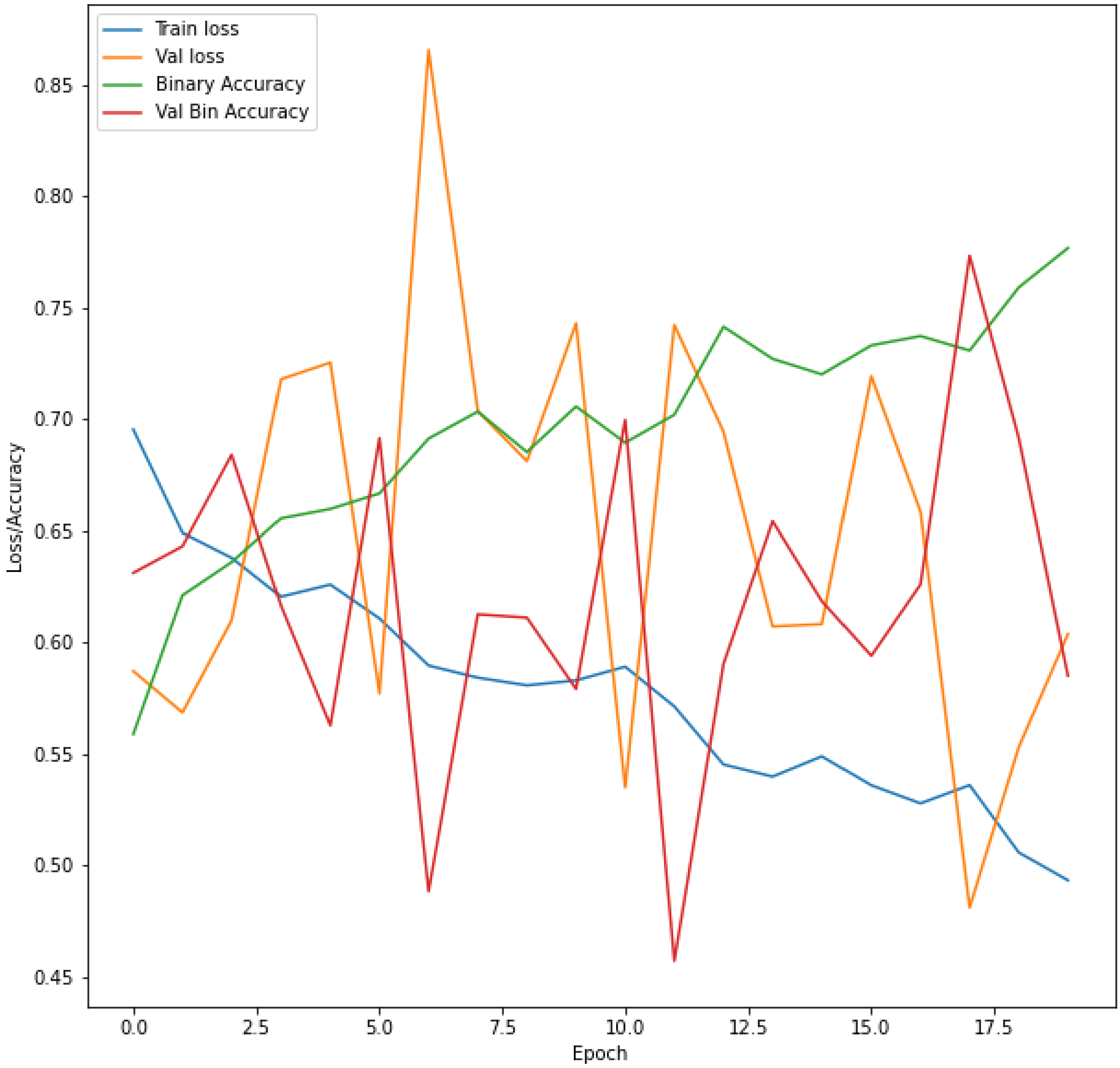
3. Algorithm Training

Parameters:

- A variety of augmentation were applied on the images during training using ImageDataGenerator from Keras library:
    - samplewise centering
    - samplewise standard normalization
    - horizontal flip
    - height shift upto 10%,
    - width shift upto 10%,
    - rotation upto 20 degrees,
    - shear upto 10%
    - zoom upto 10%
    - the size of images was also converted to 224 x 224
- Batch size of 64 images was used for training
- Adam optimizer with a learning rate of 0.0001, loss function as 'binary_crossentropy' and metrics measured was 'binary_accuracy' for training
- Layers 'input_1' to 'block5_conv2' of pre-existing architecture were frozen
- Layers 'block5_conv3' and 'block5_pool' of pre-existing architecture were fine-tuned
- Layers 'dropout_1', 'flatten_1' and 'dense_1' are added to pre-existing architecture

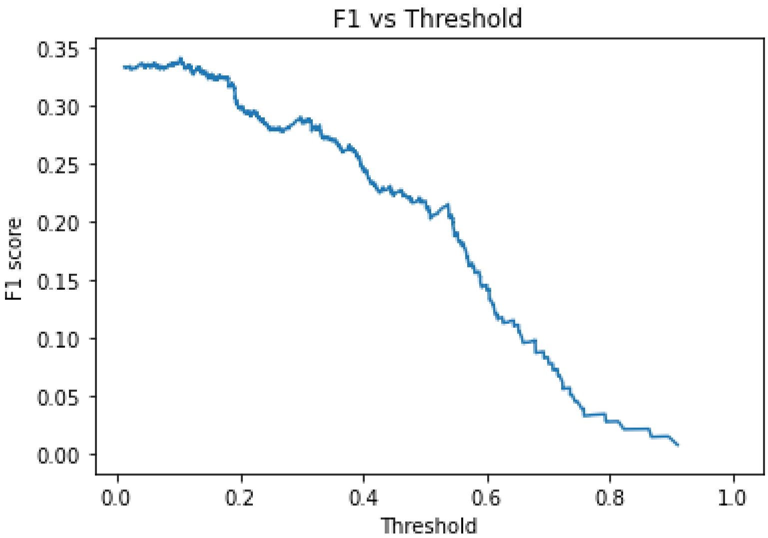
Some visualisations:

Training history plot:

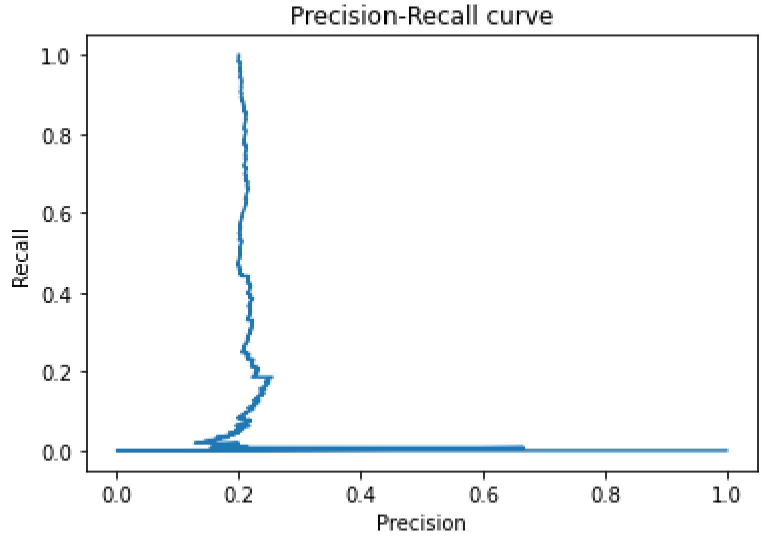


Training loss steadily reduced and stablised while validation loss initially dropped but then ossiclated in between 0.6 and 0.7.

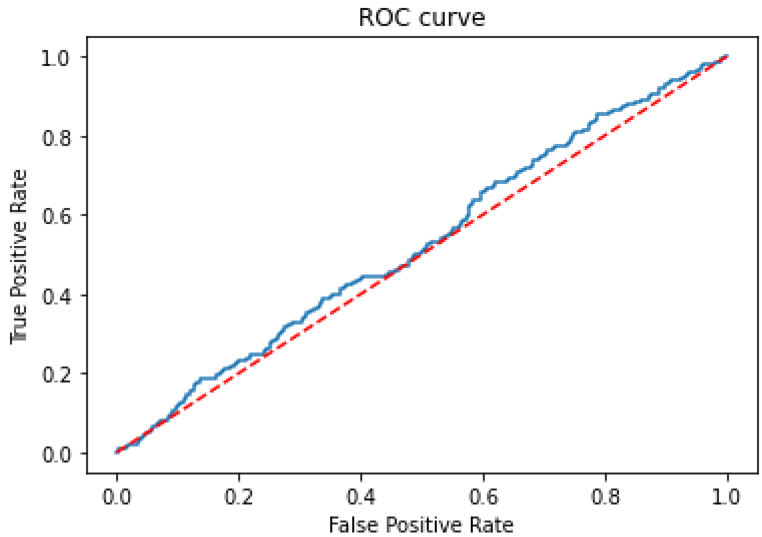
F1 score vs Threshold plot:



Precision Recall plot:



ROC curve plot:



Confusion Matrix:

229	848
228	40

Final Threshold and Explanation: The threshold chosen for classification is 0.1038 with corresponding F1 score of 0.3415, ROC AUC 0.5277, recall of 0.8513 and precision 0.2126.

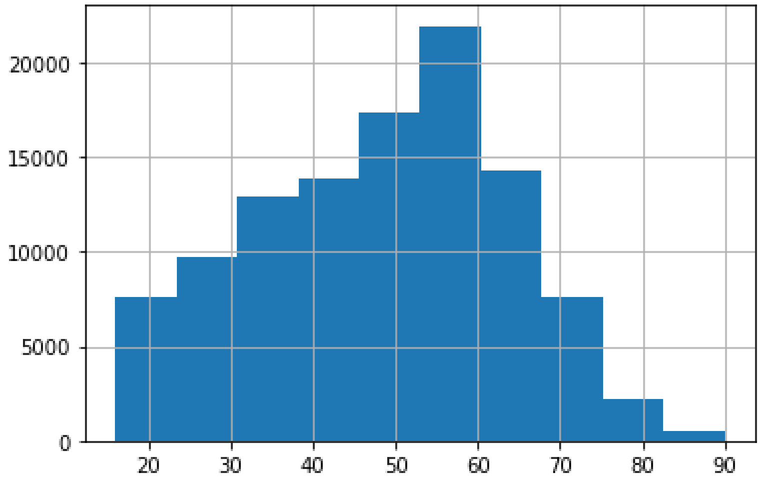
This threshold was chosen as it maximised the F1 score value.

4. Databases

Out of the full dataset consisting of 112,120 frontal-view X-ray images, 1346 images were with Pneumonia label.

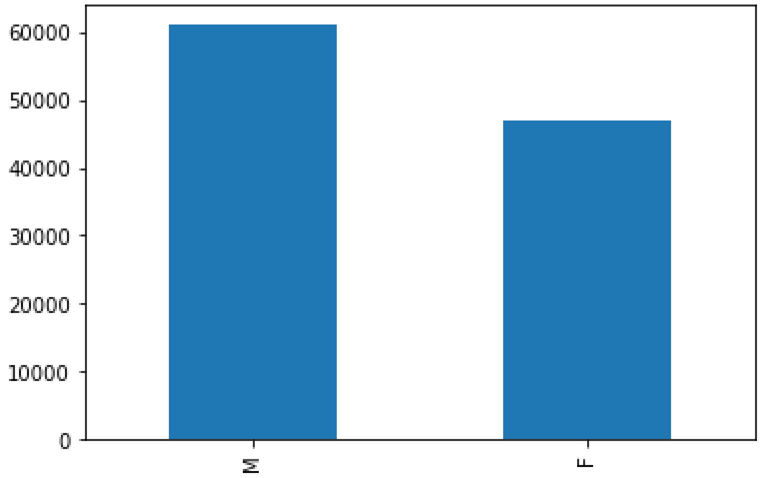
Some visualisations:

Patient demography plot:



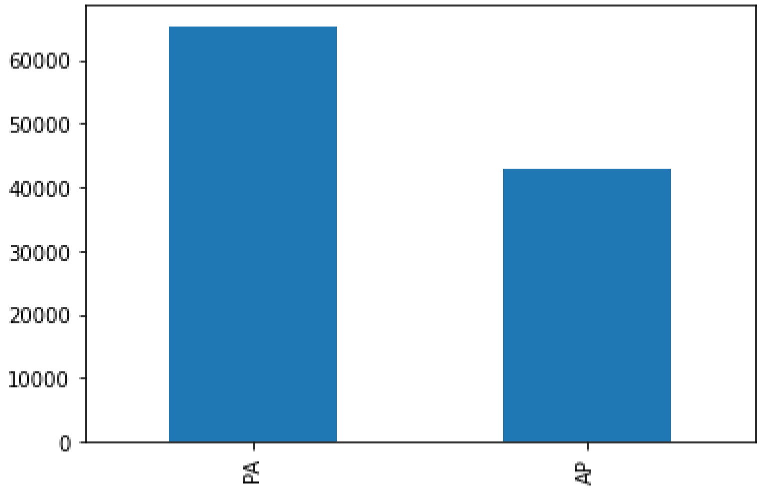
Majority of patients are aged between 16 and 75 years

Patient gender distribution plot:



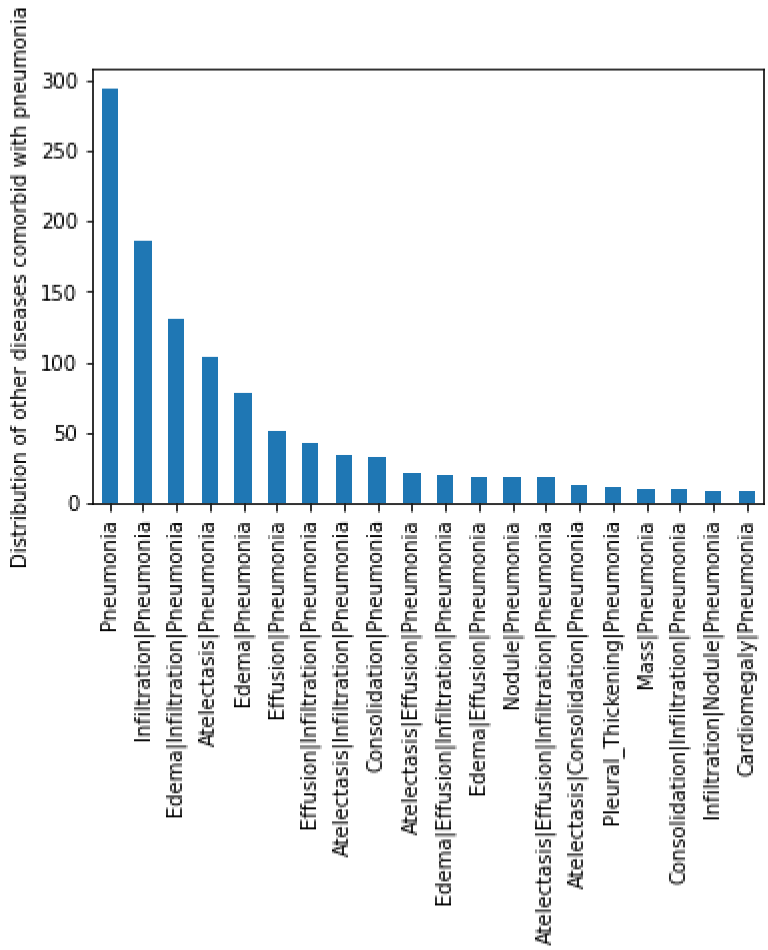
The gender distribution is slightly tilted towards Male patients.

Patient view plot:



'PA' view position is larger is number compared to 'AP' view position.

Co-occurrence frequencies of Pneumonia with other diseases and findings plot:



Number of cases comorbid with Pneumonia is highest for Infilration, followed by Edema,Atelectasis and Effusion.

The dataset was shuffled and then divided into 80-20% split between training and validation sets in a stratified manner based on Pnemonia label.

Description of Training Dataset:

The training set had xxx number of images positive for Pneumonia and xx number of images that were negative. The ratio was made 50-50% (postive: 1077, negative: 1077) to provide balance between the two classes as this was training set.

Description of Validation Dataset:

The training set had xxx number of images positive for Pneumonia and xx number of images that were negative. The ratio was 20-80% (postive: 269, negative: 1076) as the prevalance of Pneumonia in patients is assumed to be around 20% as patients are only X-rayed based on their clinical symptoms that make Pneumonia highly likely.

5. Ground Truth

The ground truth dataset was extracted from the clinical PACS database at National Institutes of Health Clinical Center and consists of ~60% of all frontal chest x-rays in the hospital. Therefore it is expected that this dataset is significantly representative to the real patient population distributions and realistic clinical diagnosis challenges.

ChestX-ray dataset comprises 112,120 frontal-view X-ray images of 30,805 unique patients with the text-mined fourteen disease image labels (where each image can have multilabels), mined from the associated radiological reports using natural language processing. Fourteen common thoracic pathologies include Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Cardiomegaly, Nodule, Mass and Hernia.

However, the limitation with this dataset is that original radiology reports (associated with these chest x-ray studies) are not meant to be publicly shared for many reasons. Hence, the image labels are NLP extracted so there would be some erroneous labels but the NLP labelling accuracy is estimated to be >90%.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: The patient population for extracting imaging data is requested to be:

- Age between 16 and 90 years
- Male and female gender distribution approximately equal
- Imaging modality of X-Ray (DX)
- Body part imaged is Chest and PatientPosition is either 'PA' or'AP'
- Prevalence of Pneumonia is expected to be around 20%
- Infiltration should ideally be excluded as comorbidities in the population

Ground Truth Acquisition Methodology:

The optimal ground truth can be created by labelling of Chest X-Ray images by a group of radiologists based on patient's medical history and lab reports. The list of 14 common diseases must be provided to them as labels and based on weighted average or voting among radiologists, the labels must be attached to the images.

This process avoids the inaccuracy potentially introduced by a NLP model for labelling images based on radiology reports.

Algorithm Performance Standard:

According to the paper 'CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning' by Pranav Rajpurkar et al., the performance standard is F1 score comparison between radiologists and model. The model in the paper achived a F1 score of 0.435, while the radiologists averaged an F1 score of 0.387. These values of F1 score can be used as a benchmark to evaluate our model.