FDA Submission

VIJAYALAKSHMI VARADHARAJAN

PNEUMO-DETECT

Algorithm Description

1. General Information:

Intended Use Statement:

To ease the work of several Radiologists to read millions of Patient's Chest X-Rays quickly and detect for the presence or absence of Pneumonia

Indications for Use:

Age Group: 1 - 95 Yrs. Old Patients' Chest X-Rays

Gender Population: Male & Female Patients' Chest X-Rays

Co-Morbidities along with Pneumonia: Edema, Infiltration, Atelectasis, Effusion, Emphysema, Pneumothorax, Mass, Hernia, Pleural Thickening, Fibrosis and Cardiomegaly

Device Limitations:

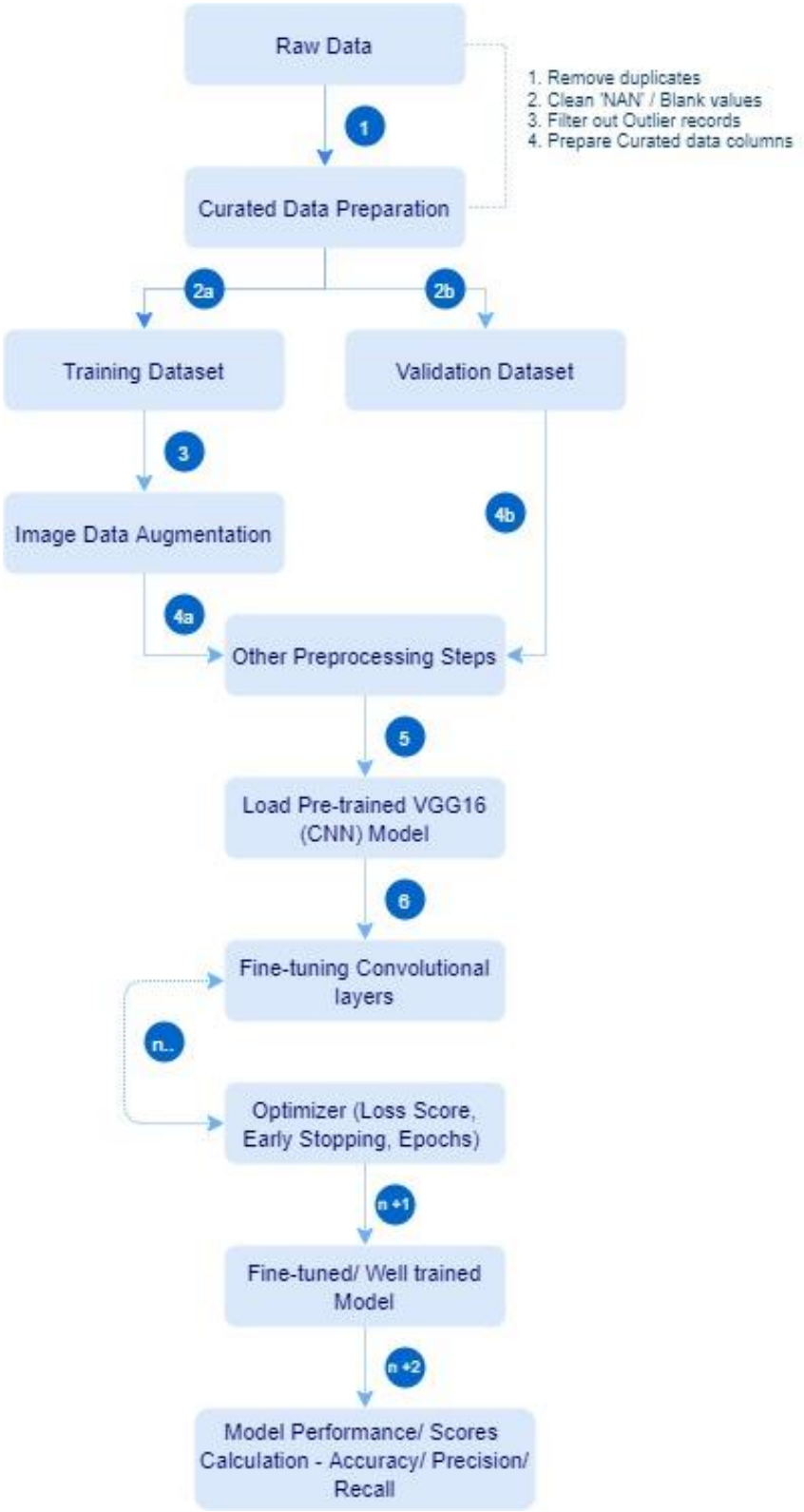
Since the device requires compute intensive operations which accelerates the performance and display results at a must faster rate, it's recommended to use the device in GPU enabled workstation.

Additionally, the images that are labelled as ‘Infiltration’ could have closer intensity distribution values to that of Pneumonia. Hence, the images labelled as ‘Infiltration’ could be mistakenly considered for ‘Pneumonia Positive’.

Clinical Impact of Performance:

The model is trained with good recall and precision scores thereby the device can be reliable to detect True Negative, True Positive, False Negative and False Positive Cases.

2. Algorithm Design and Function:



DICOM Checking Steps:

Body Part Examined : CHEST

Patient Position : AnteroPosterior (AP) & PosteroAnterior (PA) Views

Modality : Digital X-Rays (DX), Computed Tomography (CT)

Pre-processing Steps:

Image Size **:** (1, 224, 224, 3) (i.e.,) Batch, Height, Width and Channels respectively

Normalization : Data was scaled to the range of 0 to 1 i.e., [0, 1]. Rescaling argument ratio was set as 1.0/255.0

CNN Architecture:

CNN Architecture : VGG16 Model (Pre-trained Model)

CNN Layers Breakdown : Image attached below

Model: "vgg16"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, None, None, 3)]	0
block1_conv1 (Conv2D)	(None, None, None, 64)	1792
block1_conv2 (Conv2D)	(None, None, None, 64)	36928
block1_pool (MaxPooling2D)	(None, None, None, 64)	0
block2_conv1 (Conv2D)	(None, None, None, 128)	73856
block2_conv2 (Conv2D)	(None, None, None, 128)	147584
block2_pool (MaxPooling2D)	(None, None, None, 128)	0
block3_conv1 (Conv2D)	(None, None, None, 256)	295168
block3_conv2 (Conv2D)	(None, None, None, 256)	590080
block3_conv3 (Conv2D)	(None, None, None, 256)	590080
block3_pool (MaxPooling2D)	(None, None, None, 256)	0
block4_conv1 (Conv2D)	(None, None, None, 512)	1180160
block4_conv2 (Conv2D)	(None, None, None, 512)	2359808
block4_conv3 (Conv2D)	(None, None, None, 512)	2359808
block4_pool (MaxPooling2D)	(None, None, None, 512)	0
block5_conv1 (Conv2D)	(None, None, None, 512)	2359808
block5_conv2 (Conv2D)	(None, None, None, 512)	2359808
block5_conv3 (Conv2D)	(None, None, None, 512)	2359808
block5_pool (MaxPooling2D)	(None, None, None, 512)	0

Total params: 14,714,688
Trainable params: 14,714,688
Non-trainable params: 0

3. Algorithm Training:

Parameters:

Types of Augmentation used during training:

Horizontal Flip of images : Yes

Vertical Flip of images : No

Horizontal Shift Range (Width) : (+/-) 10%

Vertical Shift Range (Height) : (+/-) 10%

Rotation Range : 11%

Shear Range : 15%

Zoom Range : 20 degrees

Batch size : 32

Optimizer learning rate : 1e-4

Layers of pre-existing architecture that were frozen:

The initial 17 layers of VGG 16 CNN pre-trained model were frozen

Layers of pre-existing architecture that were fine-tuned:

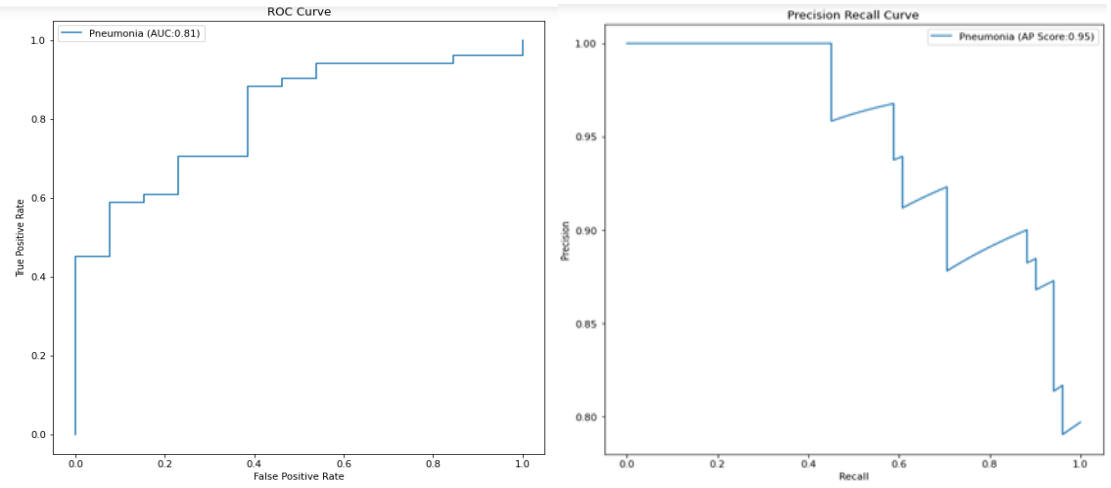
All the fully connected layers of VGG 16 CNN pre-trained model

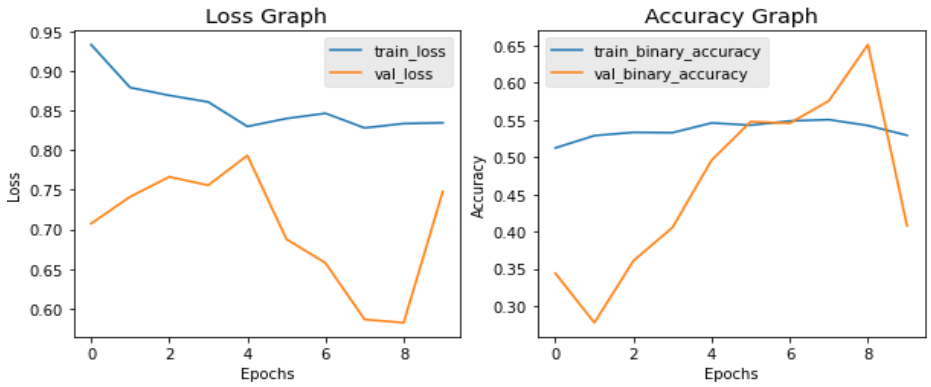
Last CNN layer

Layers added to pre-existing architecture:

Global Average Pooling 2D layer; Batch Normalization and Dense layers

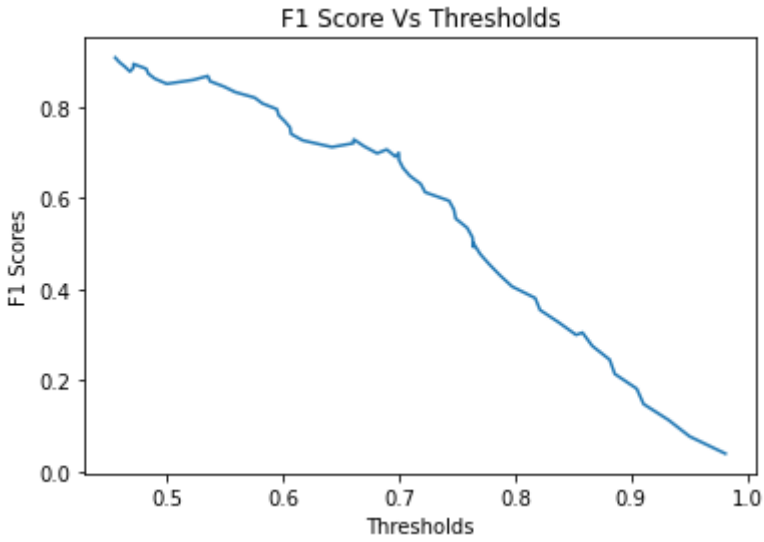
Dropout: 0.5 (To prevent model over-fitting)





Final Threshold and Explanation:

Among the two different thresholds (i.e.,) 0.45 and 0.7 calculated based on Precision and Recall scores. F1 scores were calculated for both the thresholds. F1 was greater (0.9) when the threshold was set to 0.45. Therefore, final threshold can be considered to be 0.45.



4. Databases:

Database: **NIH Dataset**

- ***Total records in Raw Dataset*** : 112,120 X-ray images
- ***Unique Patients' records*** : 30,805 Patient ID's
- ***Data Format*** : Comma delimited
- ***Metadata*** : Mentioned below
  Image Index, Finding Labels, Follow-up #, Patient ID, Patient Age, Patient Gender, View Position, Original Image Width, Original Image Height, Original Image Pixel Spacing X

dimension & Original Image Pixel Spacing y.

- **Metadata Fields Description** : Only fields which requires explanation are mentioned below ; remaining fields are self-explanatory
  'Image Index' – Image File Name
  'Finding Labels' – Disease Labels - 14 common thoracic pathologies (Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural Thickening, Pneumonia, Pneumothorax) and 'No Finding' label indicates that no disease was identified for the corresponding image

Description of Training Dataset:

- 50-50 (%) equal splitting of Pneumonia & Non-Pneumonia Cases respectively
- ***Total records in Training dataset*** : 2290 Images

Description of Validation Dataset:

- 80-20 (%) splitting of Non-Pneumonia & Pneumonia Cases respectively
- ***Total records in Testing dataset:*** 1430

5. Ground Truth:

The disease labels were created using Natural Language Processing (NLP) to mine the associated radiological reports.

The biggest limitation of this NIH 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%.

6. FDA Validation Plan:

Patient Population Description for FDA Validation Dataset:

Age Group	: 1 - 95 Yrs. Old Patients' Chest X-Rays
Gender Population	: Male & Female Patients' Chest X-Rays
Data Split Ratio	**:** 80-20 (%) splitting of Non-Pneumonia & Pneumonia Cases respectively
Body Part Examined	: CHEST
Patient Position	: AnteroPosterior (AP) & PosteroAnterior (PA) Views
Modality	: Digital X-Rays (DX), Computed Tomography (CT)

Comorbidity conditions for Pneumonia : Patients with Infiltration, Edema, Effusion, Consolidation and Atelectasis must be included for Pneumonia detection.

Ground Truth Acquisition Methodology:

The image labels were NLP-extracted with NLP labelling accuracy is estimated to be >90%.

Algorithm Performance Standard:

As per the definition, F1 score is the weighted average of Precision – Recall Scores. F1 scores helps us to interpret the uneven class distribution. The F1 score achieved from the device 'Pneumo-Detect' with the underlying VGG16 fine-tuned model is 0.9 which is far higher when compared to Radiologist's average highest F1 score of 0.4.