FDA SUBMISSION

Name: Aryan S Mishra
Name of the Device: Pneumonia X-ray

Algorithm Description

General Information

Intended Use Statement: The recognition of Pneumonia using a Chest X-Ray is a software-based device that allows the review, analysis and interpretation of chest X-ray images. It is not to be used as a primary interpretation. The software provides the classification of Pneumonia, numerical analysis and substance indication. The user can review, verify and correct the results of the system and generate a report based on the conclusion.

Indications for Use: The Pneumonia X-ray is a software designed to aid the clinical assessment of the newborn to adult chest X-ray cases with featured suggestive of pneumonia in the medical environment. pneumonia-ray analyzes cases using an artificial algorithm to identify suspected findings of pneumonia. It makes case level output available in a PACS system for worklist prioritization. pneumonia-ray is not intended to direct attention to a specific portion or anomalies of an image. Its results are not intended to be used on a standalone basis for clinical decision making nor is it intended to rule out pneumonia or otherwise preclude clinical assessment of X-ray cases.

Device Limitations: Pneumonia X-ray software is designed for analysis of X-ray images captured in 2-D format only and does not support any other form of medical images like CT or MRI scan images etc. pneumonia-ray performs a binary classification of the prevalence of Pneumonia in a given X-ray and does not perform segmentation or localization of Pneumonia. The software is designed by training data of a newborn baby to adults up to 100 years of age both male and female.

Clinical Impact of Performance: Pneumonia X-ray assists medical professionals in featured suggestive of pneumonia in the medical environment there by reducing manual effort and human error in analyzing an X-ray image but does not replace the activities of a physician. A prediction of false positive or false negative may affect the treatment plan for a patient, hence it is recommended to have a strong review mechanism by a qualified physician for every case

Algorithm

- Exploratory Data Analysis

From the data the following exploratory analysis will be done
* The patient demographic data such as gender, age, patient position, etc. (as it is available)
* The x-ray views taken (i.e. view position)
* The number of cases including:
  * number of pneumonia cases,
  * number of non-pneumonia cases
* The distribution of other diseases that are comorbid with pneumonia
* Number of disease per patient
* Pixel-level assessments of the imaging data for healthy & disease states of interest (e.g. histograms of intensity values) and compare distributions across diseases.

- Building and Training the Model

Training and validating Datasets

From the findings in the EDA component of this project, the following are taken into consideration

* Distribution of diseases other than pneumonia that are present in both datasets
* Demographic information, image view positions, and number of images per patient in each set
* Distribution of pneumonia-positive and pneumonia-negative cases in each dataset

- Model Architecture

In this project, will fine-tune an existing CNN architecture to classify x-rays images for the presence of pneumonia. There is no required architecture required for this project, a reasonable choice would be using the VGG16 architecture with weights trained on the ImageNet dataset.
Fine-tuning can be performed by freezing the chosen pre-built network and adding several new layers to the end to train, or by doing this in combination with selectively freezing and training some layers of the pre-trained network.

- Image Pre-Processing and Augmentation

It serves the purpose of conforming model's architecture and/or for the purposes of augmenting your training dataset for increasing your model performance. When performing image augmentation, be sure to think about augmentation parameters that reflect real-world differences that may be seen in chest X-rays.

- Training

In training your model, there are many parameters that can be tweaked to improve performance including:

* Image augmentation parameters
* Training batch size
* Training learning rate
* Inclusion and parameters of specific layers in your model

- Performance Assessment

The models performance is monitored over subsequence training epochs.

- Clinical Workflow Integration

The imaging data provided to was transformed from DICOM format into .png to help aid in the image pre-processing and model training steps of this project.

For this project a DICOM wrapper is created that takes in a standard DICOM file and outputs data in the format accepted by the model.

* Proper image acquisition type (i.e. X-ray)
* Proper image acquisition orientation (i.e. those present in your training data)
* Proper body part in acquisition

DICOM Checking Steps:

DICOM header provide all the attributes except for the pixel data of the image and the image file provide pixel data representing an actual image. The first step is to pre-extract key attributes from DICOM headers into a data frame using the Python package pyidcom to optimize image processing workflow and algorithm training. Since PneumoniaX-ray is specifically designed for deducting pneumonia in 2D chest X-ray, we ensure to check image position is PosteriorAnterior (PA) or Antero-Posterior(AP), the image type is Digital radiography (DX), and the body part is of Chest in the given image. Any image not meeting the criteria is not processed by the software.

Pre-processing Steps: The NIH dataset is available in a CSV file “Data_Entry_2017.csv”. The following steps are followed for pre-processing

1) Load the CSV file data in a data frame using Python Pandas library

2) ‘Finding Labels’ column stores disease findings from the X-ray image. In many cases, more than one disease is prevalent in an X-ray image and each disease is separated by ‘|’ in the Finding Labels column. As good additional columns for each disease (14 types in this dataset) are created and have a binary representation for presence and no presence of pneumonia for easy data manipulation and analysis

3) X-ray image files are stored in a separate directory. Use ‘os’ and ‘glob’ modules to read the image file path and stored it in a column against each image for easy manipulation

CNN Architecture: Pre trained VGG16 architecture with weights trained on the ImageNet dataset with added dropouts and dense layer

3. Algorithm Training Parameters: For PneumoniaX-ray algorithm we have used a pretrained model called “VGG16” and fined tuned the model based on the dataset. Below is the parameters for “VGG16”

Model: "vgg16"

Layer (type)	Output Shape	Param #
input_2 (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
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312
predictions (Dense)	(None, 1000)	4097000

===

Total params: 138,357,544
Trainable params: 138,357,544
Non-trainable params: 0

Types of augmentation used during training : All the images are augmented to have uniformity across all the images in the training set. Hence the images are rescaled to 1./255 - every pixel value from range [0,255] -> [0,1]. Flipping the image horizontally is set to true and vertical flip is set the false.

Batch size : Batch Size of 64 is used

Optimizer learning rate : In PneumoniaX-ray 'Adam' optimizer is used. Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments. Learning rate of 1e-4(0.0001) is used

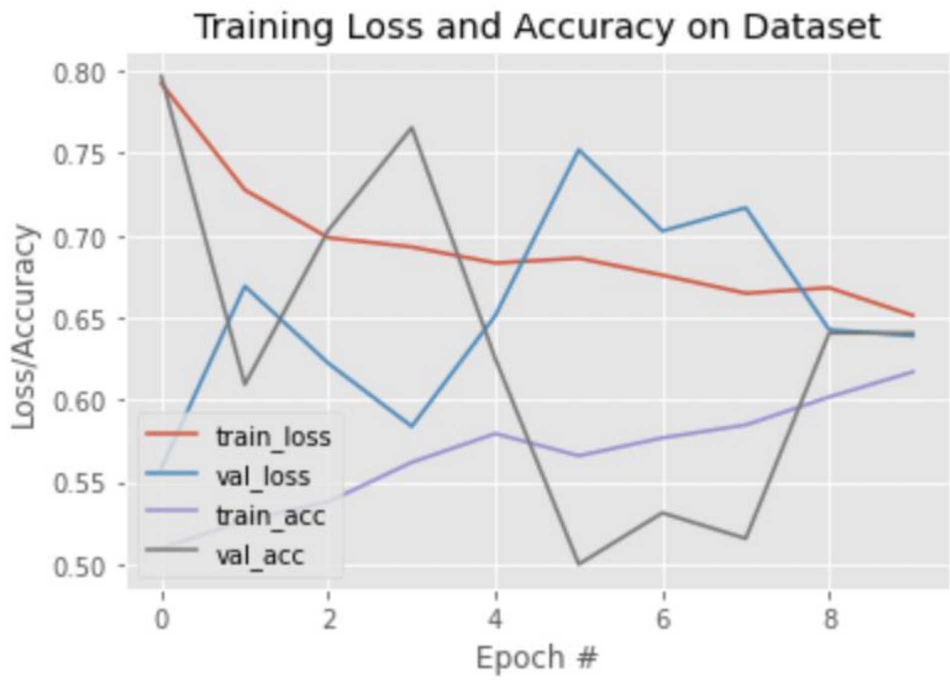
Layers of pre-existing architecture that were frozen : Please refer to the flow chart in CNN Architecture. VGG16 architecture with weights trained on the ImageNet dataset, Conv Block #1 to #5 containing is frozen

Layers of pre-existing architecture that were fine-tuned : Loaded the pretrained weights and train the complete network with a smaller learning rate of 1e-4(0.0001). This results in very good accuracy with even small datasets

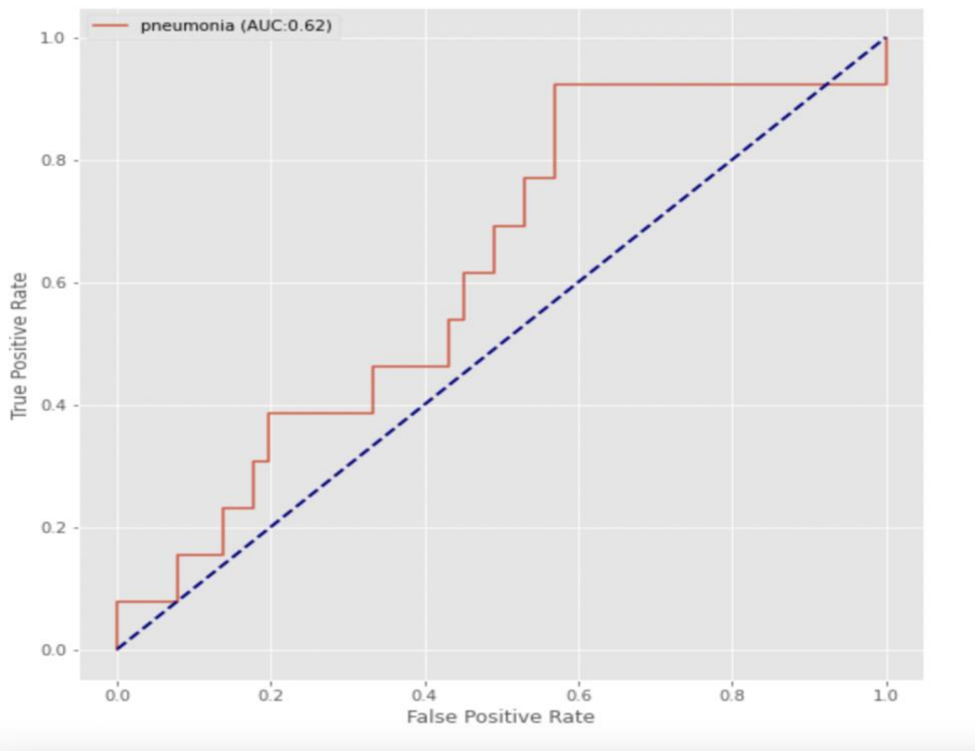
Layers added to pre-existing architecture : Added a classifier on top of the convolutional base by adding a fully connected layer followed by a softmax layer with 4 outputs.

Algorithm training performance visualization :

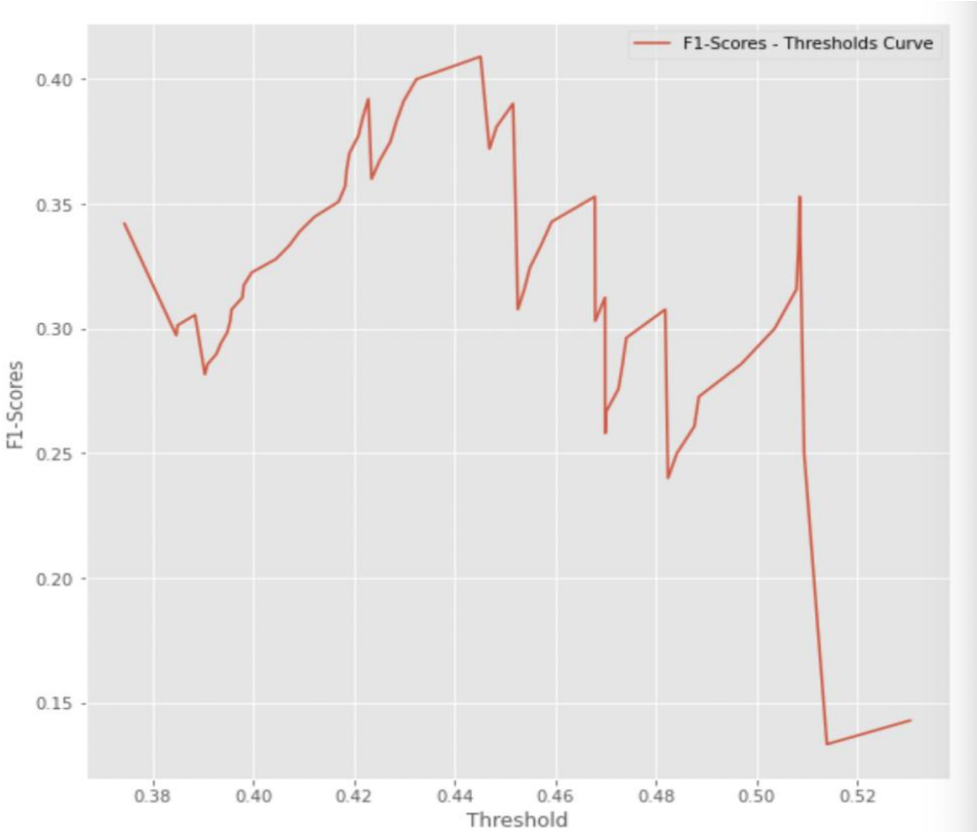
```
plot_history(history)
```



P-R curve



F1-Scores – Thresholds Curve



Final Threshold and Explanation: Final Threshold of 0.62786263 at 80% Precision is considered with F1-score of 0.14285. Test accuracy is at 67.18%, AUC 0.64 and AP score of 0.3. Looking at the above graph threshold of 0.445 at F1-Score of 0.42 seems to be optimum, however given the 6 test images, where in predicted threshold is 0.4775539, final threshold of 0.6 2786263 gives accurate result on test images

4. Dataset

The dataset was curated by NIH specifically to address the problem of a lack of large x-ray datasets with ground truth labels to be used in the creation of disease detection algorithms. 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:

- Atelectasis
- Consolidation
- Infiltration
- Pneumothorax
- Edema
- Emphysema
- Fibrosis

- Effusion
- Pneumonia
- Pleural thickening
- Cardiomegaly
- Nodule
- Mass
- Hernia

One can identify three limitations of this comparison. First, only frontal radiographs were presented to the radiologists and model during diagnosis, but it has been shown that up to 15% of accurate diagnoses require the lateral view (Raoof et al., 2012); thus expect that this setup provides a conservative estimate of performance. Third, neither the model nor the radiologists were permitted to use patient history, which has been shown to decrease radiologist diagnostic performance in interpreting chest radiographs (Berbaum et al., 1985; Potchen et al., 1979); for example, given a pulmonary abnormality with a history of fever and cough, pneumonia would be appropriate rather than less specific terms such as infiltration or consolidation)

An ideal database should provide the different angles of chest X-ray along with medical history of patient for better prediction of phenomena

Description of Training Dataset: NIH X-ray image dataset which also contains additional information like patient details and annotated information like disease prevalence is split into Training Dataset and Validation dataset in the ratio of 80:20 using python library skl.train_test_split. In the training dataset it is ensured that there is balanced dataset of images with pneumonia and images without pneumonia

Description of Validation Dataset: Validation dataset represent the 20% of the overall dataset for prevalence of pneumonia with imbalanced mix of images with pneumonia and images without pneumonia. It is ensured that images used in Training dataset is not used in validation dataset.

5. Ground Truth

Ground Truth is acquired from NIH prepared datasets of 112,120 X-ray images with 14 disease labels from 30,805 unique patients. The image labels were extracted using NLP 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: The dataset contains 30,805 unique patients in the age group of new born to 100 years of age with an approx. mix of 44% and 56% of female and male population

Ground Truth Acquisition Methodology: Ground truth acquisition methodology can be categorised into two major category

Gold standard

The gold standard for a particular type of data refers to the method that detects disease with the highest sensitivity and accuracy. Any new method that is developed can be compared to this to determine its performance. Typical sources of ground truth for Pneumonia are :

- Chest X-ray to look for inflammation in lungs
- Blood test to check white blood cell count
- Sputum tests (using a microscope to look at the gunk you cough up)
- A pulse oximetry test, which measures the oxygen in your blood

Algorithm Performance Standard: The PneumoniaX-ray has been evaluated and verified in accordance with software specifications and applicable performance standards through Software Development and Validation & Verification Process to ensure performance according to specifications, User Requirements and Federal Regulations and Guidance documents, "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices". The performance of the PneumoniaX-ray device has been validated in a pivotal performance study that was carried out in simulated synthetic work-flow. Below are some of the key performance metrics

TRAIN METRIC ----------------------
Train acc: 61.7%
TEST METRICS ----------------------
Accuracy: 79.6875%
True Negative: 51
True Positive: 0
False Negative: 13
False Positive: 0

Sensitivity: 0.0

specificity: 1.0

Confusion Matrix: [[51 0]
 [13 0]]

Threshold where Precision is 0.8--------------------
Precision is: 1.0
Recall is: 0.07692307692307693
Threshold is: 0.4328137
F1 Score is: 0.14285714285714288

Threshold where Recall is 0.8--------------------
Precision is: 0.2564102564102564
Recall is: 0.7692307692307693
Threshold is: 0.34997073
F1 Score is: 0.38461538461538464