# FDA Submission

\*\*Kunnipa Prae-arporn\*\*

\*\*Algorithm for Pneumonia Classification\*\*

##Algorithm Description

### 1. General Information

\*\*Intended Use Statement: \*\*

The algorithm is used to detect Pneumonia, for improving the radiologist's workflow and acts as an assistance when making the decision.

\*\*Indications for Use: \*\*

* The algorithm can only be deployed with Chest x-ray images in DICOM format.
* The x-ray chest image must be taken in AP/PA position only
* The algorithm is compatible for both men and women from 1 to 95 years old.

\*\*Device Limitations: \*\*

* GPU is required for training the mode
* dataset needs to be 2D Xray image only

\*\*Clinical Impact of Performance: \*\*

* The model has a tradeoff between precision and recall. If the algorithm has a high recall, it indicates that the performance for screening studies and worklist prioritization will be better, since it does not take the FP into account. Whereas, a high precision means that the algorithm has more confidence in a positive cases (Positive Predictive Value), which is useful for confirming the diagnosis.

### 2. Algorithm Design and Function

Model Definition

Model Training

Validation

Inference

Data Loader

Report

Preprocessing

\*\* DICOM Checking steps: \*\*

* The dataset was read using ‘dcmread’ function and was arranged into a pixel array.
* The dataset was resized so that it matches the size of the first layer of the model.
* “CHEST” is the part of the body that is examined, and the patient position is either “PA” or “AP” or “PA”

\*\* Preprocessing Steps: \*\*

* Image normalisation is applied to the dataset
* Image is resized to (224,224,3) where 3 represents 3 colour channels.
* Augmentation is performed to the image to prevent model overfitting:

my\_idg = ImageDataGenerator(rescale = 1 / 255.0,

horizontal\_flip = True,

vertical\_flip = False,

height\_shift\_range = 0.1,

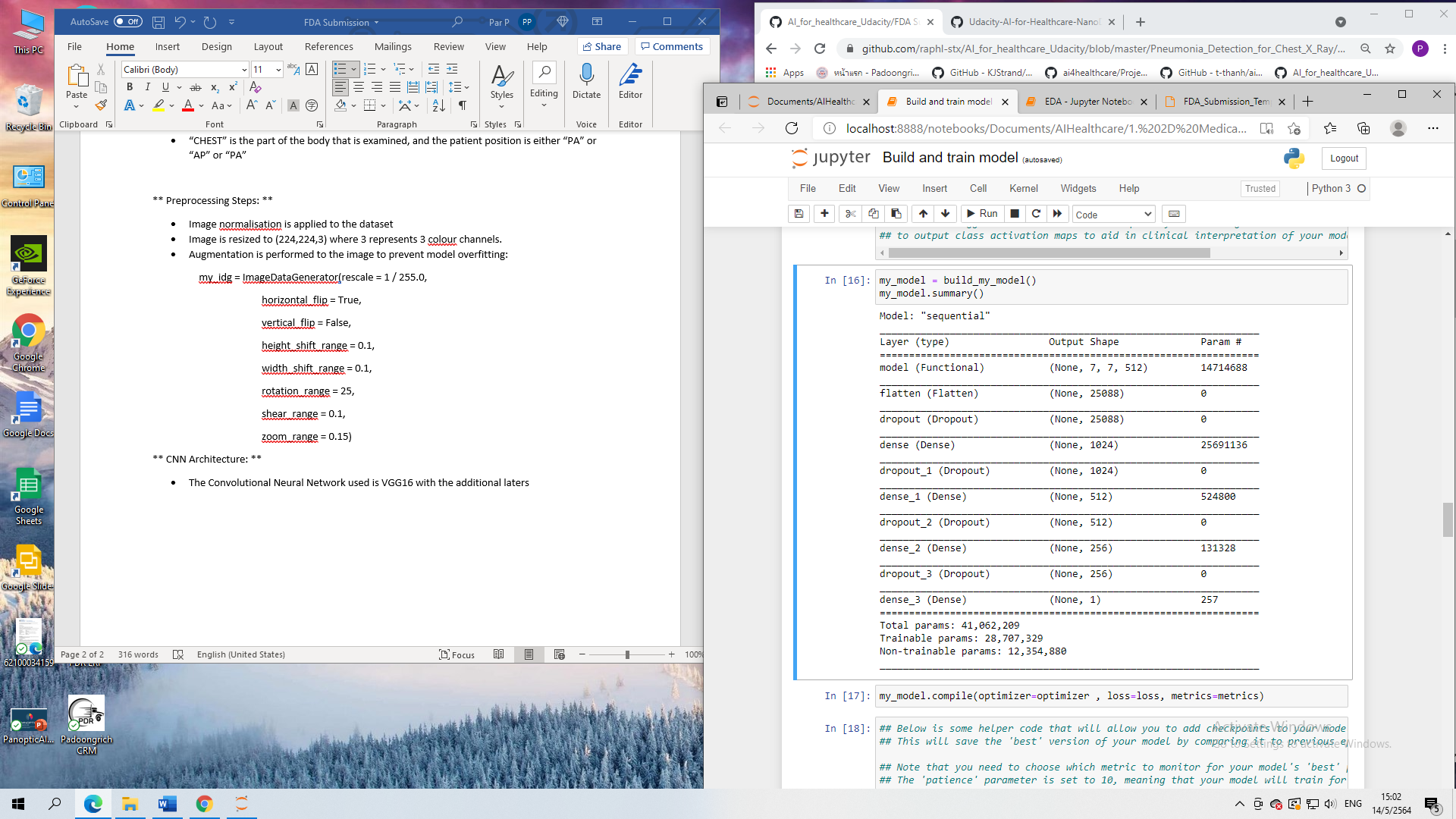
width\_shift\_range = 0.1,

rotation\_range = 25,

shear\_range = 0.1,

zoom\_range = 0.15)

\*\* CNN Architecture: \*\*

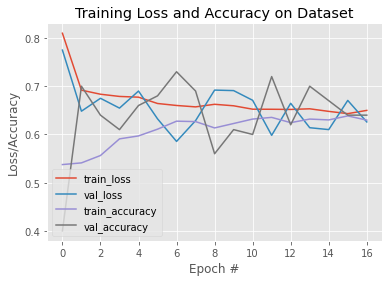
* The Convolutional Neural Network used is VGG16 with the additional layers:
* The activation function used in each Dense layer is ‘relu’ with a dropout rate of 20%. The activation function in the last fully connected layer is the ‘ sigmoid’ function.

### 3. Algorithm Training

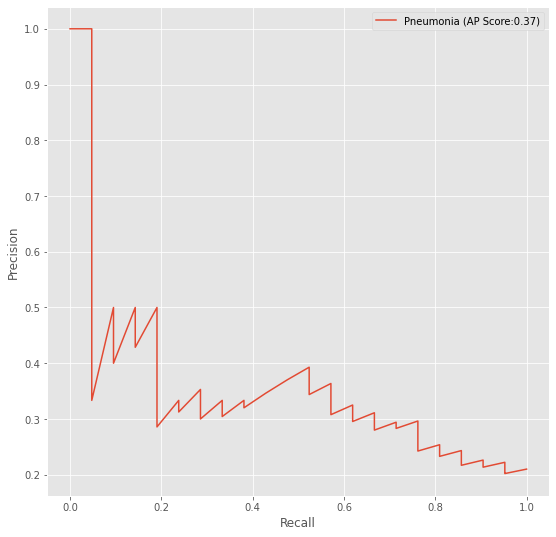
\*\* Parameters: \*\*

* Batch size = 32
* Optimiser = Adam
* Learning rate = 1e-3
* Epochs = 30
* Loss = binary cross entropy
* Augmentation parameters:
  + Only horizontal flips, no vertical flips
  + Height shift range is set to 0.1
  + Width shift range is set to 0.1
  + Shear range is set to 0.1
  + Zoom range is set to 0.15
* The ‘block5\_pool’ of VGG model (pre-existing architecture) was frozen
* Output from VGG model were not fine-tuned
* 1 Flatten, 4 Dropout, and 4 Dense layers were added with relu activation function in the first three Dense layer and sigmoid activation function in the last layer

<< Insert algorithm training performance visualization >>



<< Insert P-R curve >>



\*\*Final Threshold and Explanation: \*\*

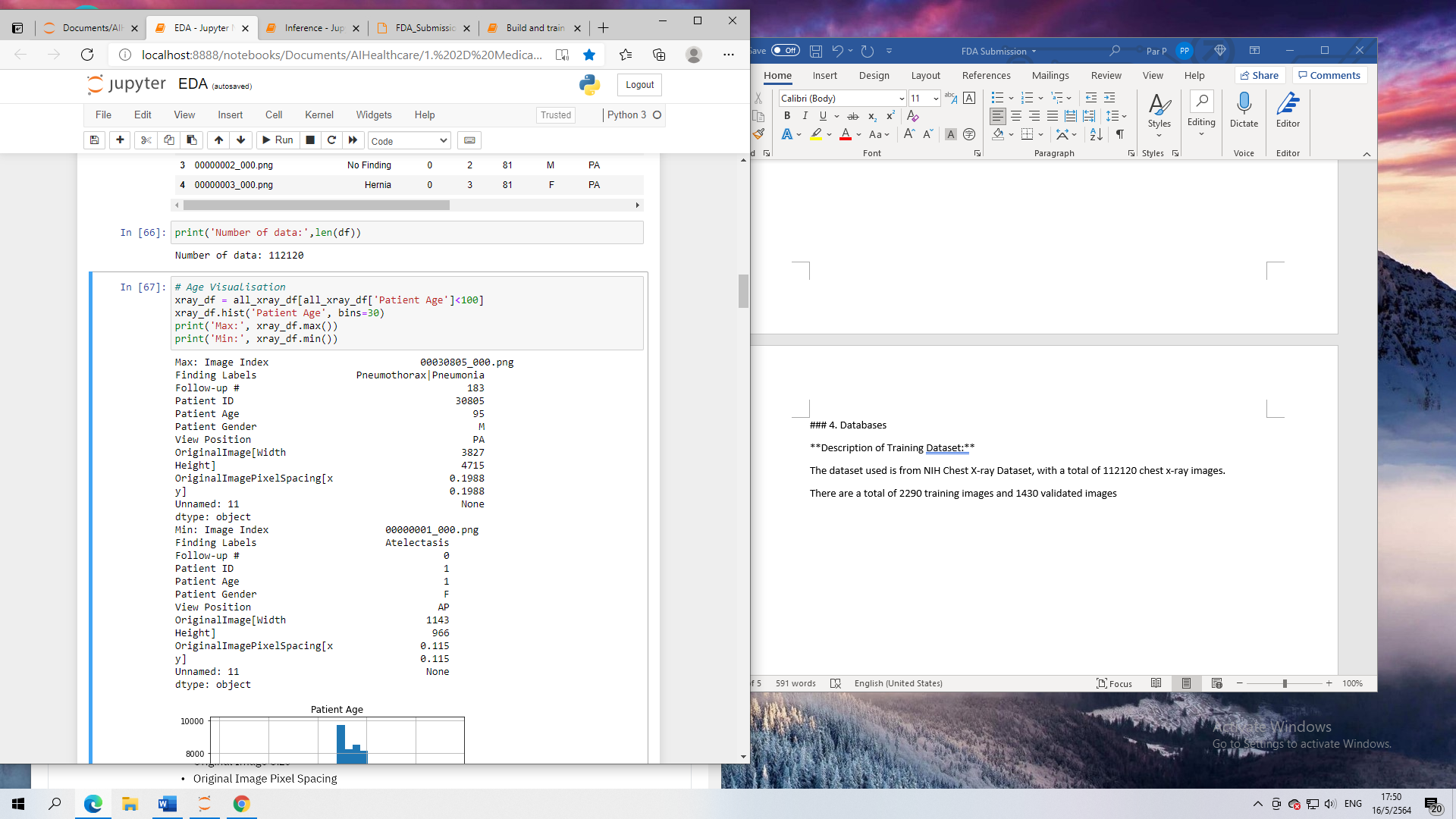
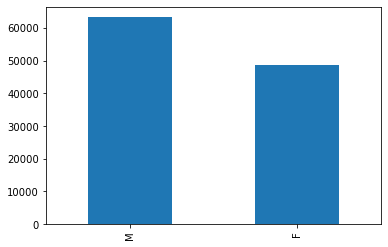
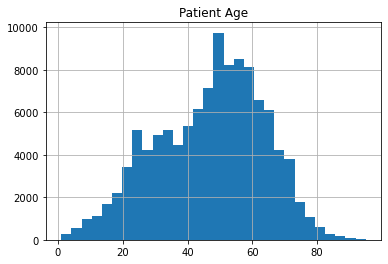
* Precision: 0.393
* Recall: 0.524
* Threshold: 0.434
* F1 Score: 0.449

The final threshold for the model is 0.434 which corresponds to the best output F1 Score pf 0.449. According to F1 Score if the prediction is above 0.43, then the output is pneumonia. Here, we can see that recall is higher than precision. As a result, classification algorithm for pneumonia is good when recall is higher because high recall indicates a more accurate screening studies since the number of TN is high.

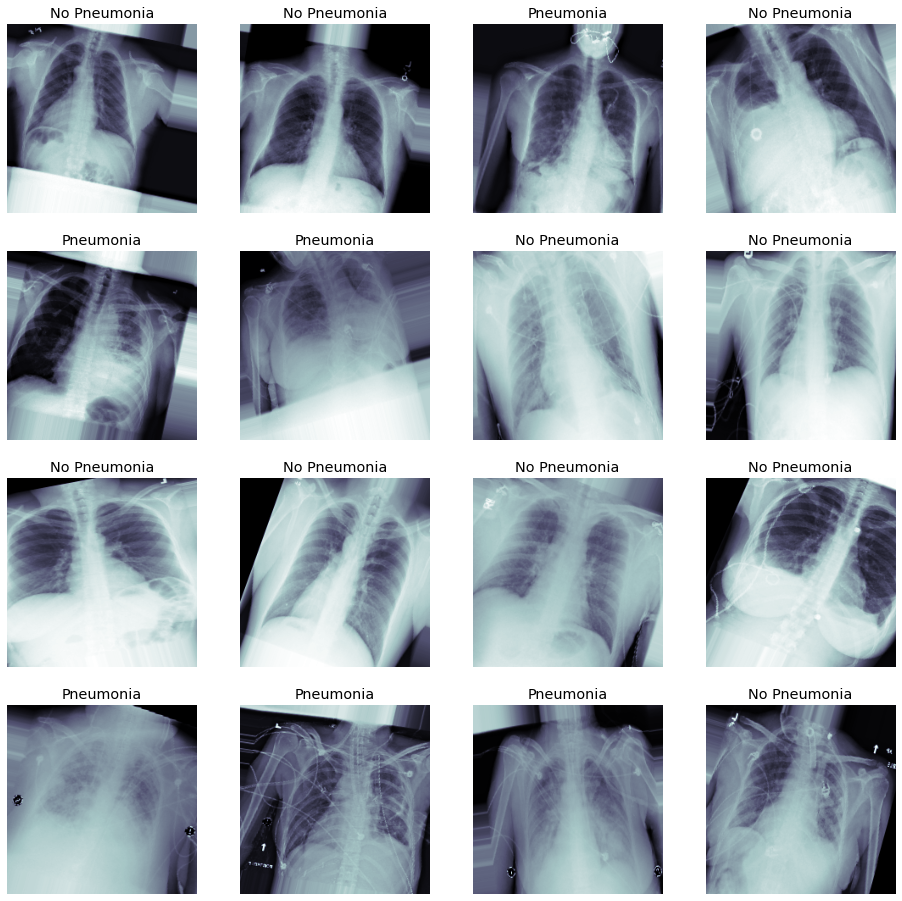
### 4. Databases

\*\*Description of Training Dataset: \*\*

The dataset used is from NIH Chest X-ray Dataset, with a total of 112120 chest x-ray images. Each of the DICOM images possess these information:

* Image Index
* Finding Labels
* Follow-up #
* Patient ID
* Patient Age
* Patient Gender
* View Position
* Original Image Size
* Original Image Pixel Spacing

\*\*Description of Training Dataset: \*\*

There are a total of 2290 training images, with a balanced number of positive and negative Pneumonia (50|50). Images in this dataset has the same relevant patient information as stated in above.

\*\*Description of Validation Dataset: \*\*

There is a total of 1430 images from validation dataset. Where, the image also displays the same relevant patient information as the training set. The dataset is split into 20% with Pneumonia positive and 80% with Pneumonia negative

### 5. Ground Truth

* A Total of 112120 x-ray images in a DICOM format with 14 labels of common thoracic pathologies, It was given that the labels are expected to be more than 90 percent accurate since the labels were obtained using Natural Language Processing:
  + Atelextasis
  + Cardiomegaly
  + Consolidation
  + Edema
  + Effusion
  + Emphysema
  + Fibrosis
  + Hernia
  + Infiltration
  + Mass
  + Nodule
  + Pleural Thickening
  + Pneumonia
  + Pneumothorax

### 6. FDA Validation Plan

\*\*Patient Population Description for FDA Validation Dataset:\*\*

* Only applicable for men and women
* Age 1 to 95
* Dataset in DICOM format
* Image must be in AP or Pa position

\*\*Ground Truth Acquisition Methodology:\*\*

A silver standard with 3 radiologists validating the final ground truth through voting

\*\*Algorithm Performance Standard:\*\*

The metric used is in the training is the binary cross entropy loss. The average F1 score for radiologists is 0.38. Wheres, the F1 score achieved from this model is 0.45 which exceeds the radiologist threshold.