FDA_submission

FDA Submission

Your Name: Luqman Awoniyi

Name of your Device: Pneumonia detector

Algorithm Description

1. General Information

Intended Use Statement: This algorithm employ Convolutional Neural Network (CNN) model to detect the presence of pneumonia in X-ray images. The aim is to assist radiogist in reviewing X-ray images.

Indications for Use: This algorithm may be used for patient, both male and female withing the age bracket of 1 to 95 whose medical presentation sugested pneumonia.

Device Limitations: * Infiltration and effusion can confuse this model and lead to false results. * This algoritm run effeicient on computers with minimum GPU and RAM requirements.

Clinical Impact of Performance:

2. Algorithm Design and Function

« Insert Algorithm Flowchart »

DICOM Checking Steps: The DICOM files are initailly checked for the following: * Is Patient position in PA or AP position? * Is BodyPartExamined==‘CHEST’? If any of these is no then the DICOM file not opened.

Preprocessing Steps: * Normalization * Resizing of images to 512 by 512 pixel

CNN Architecture: This model was built by Fine-tuning the VGG16 Model with ImageNet weights. It uses part of VGG16 layers untill the block5_pool. These following layers were added to the frozen VGG16 layers:

my_model.add(GlobalAveragePooling2D())

my_model.add(Dense(256, activation=‘relu’))

my_model.add(BatchNormalization())

my_model.add(Dropout(0.4))

my_model.add(Dense(64, activation=‘relu’))

my_model.add(BatchNormalization())

my_model.add(Dropout(0.3))

my_model.add(Dense(32, activation='relu'))

my_model.add(BatchNormalization())

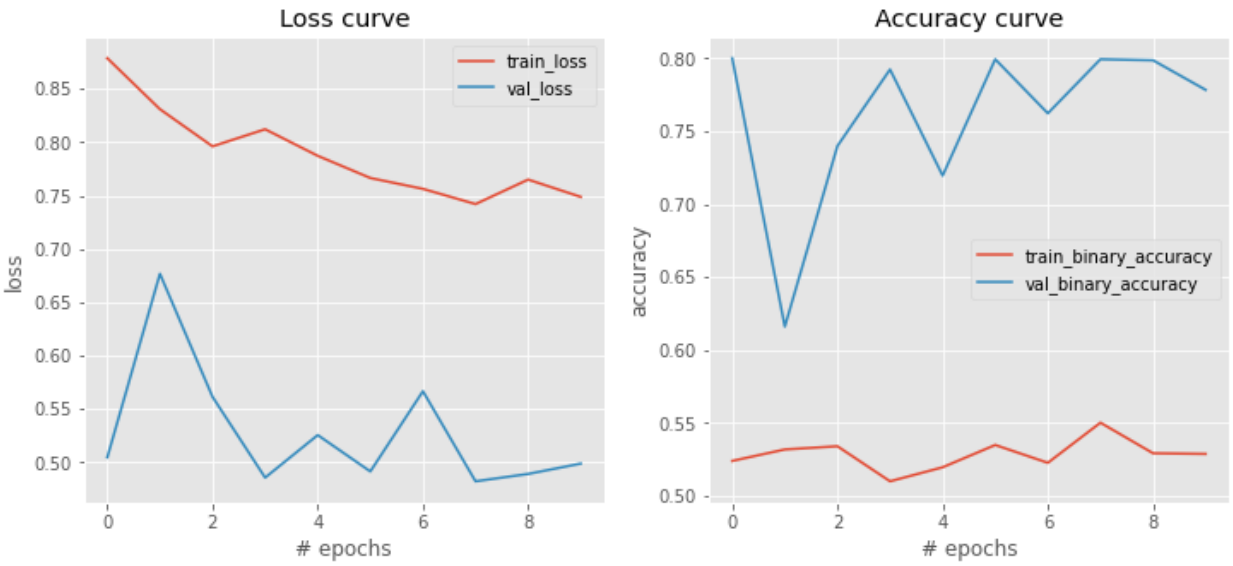
my_model.add(Dropout(0.3))

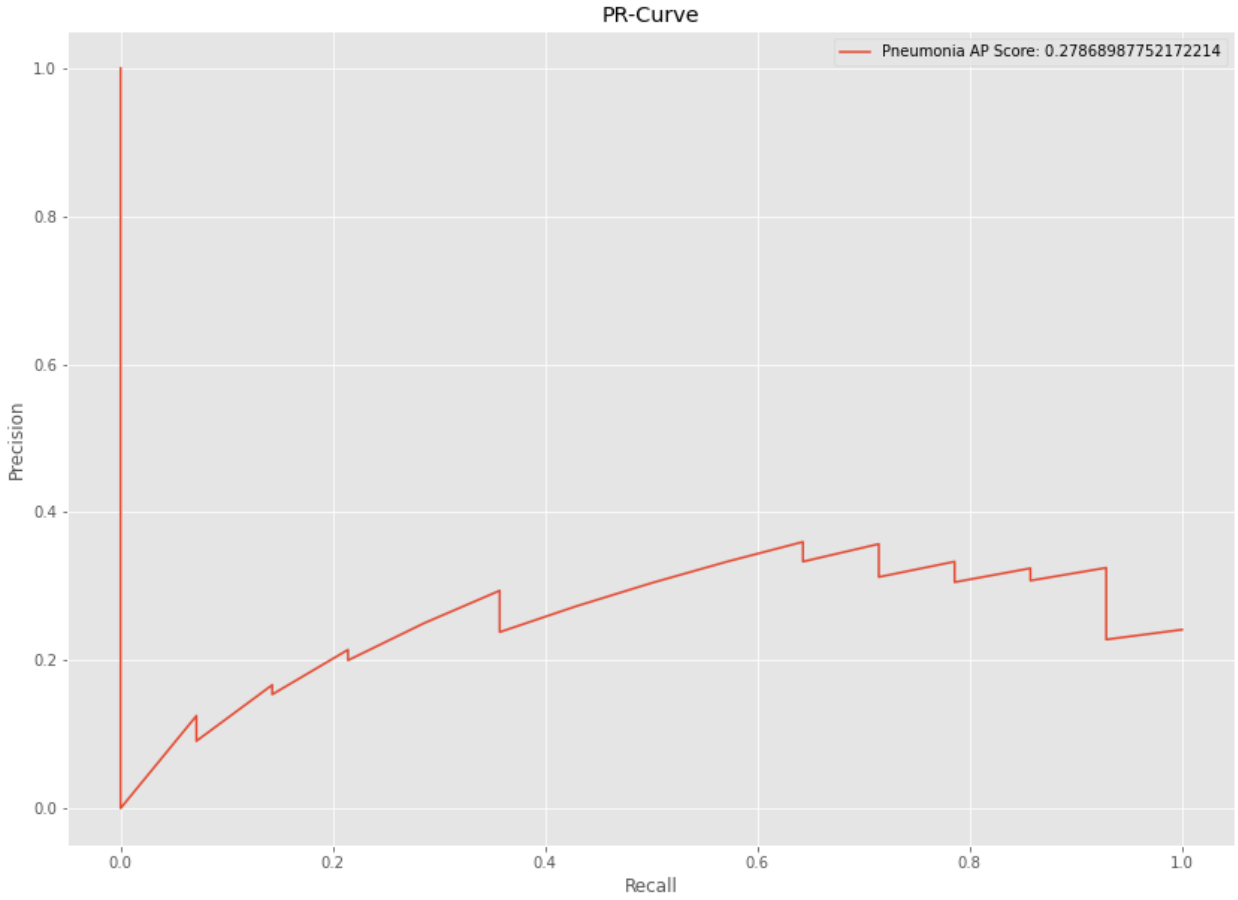
my_model.add(Dense(1, activation='sigmoid'))

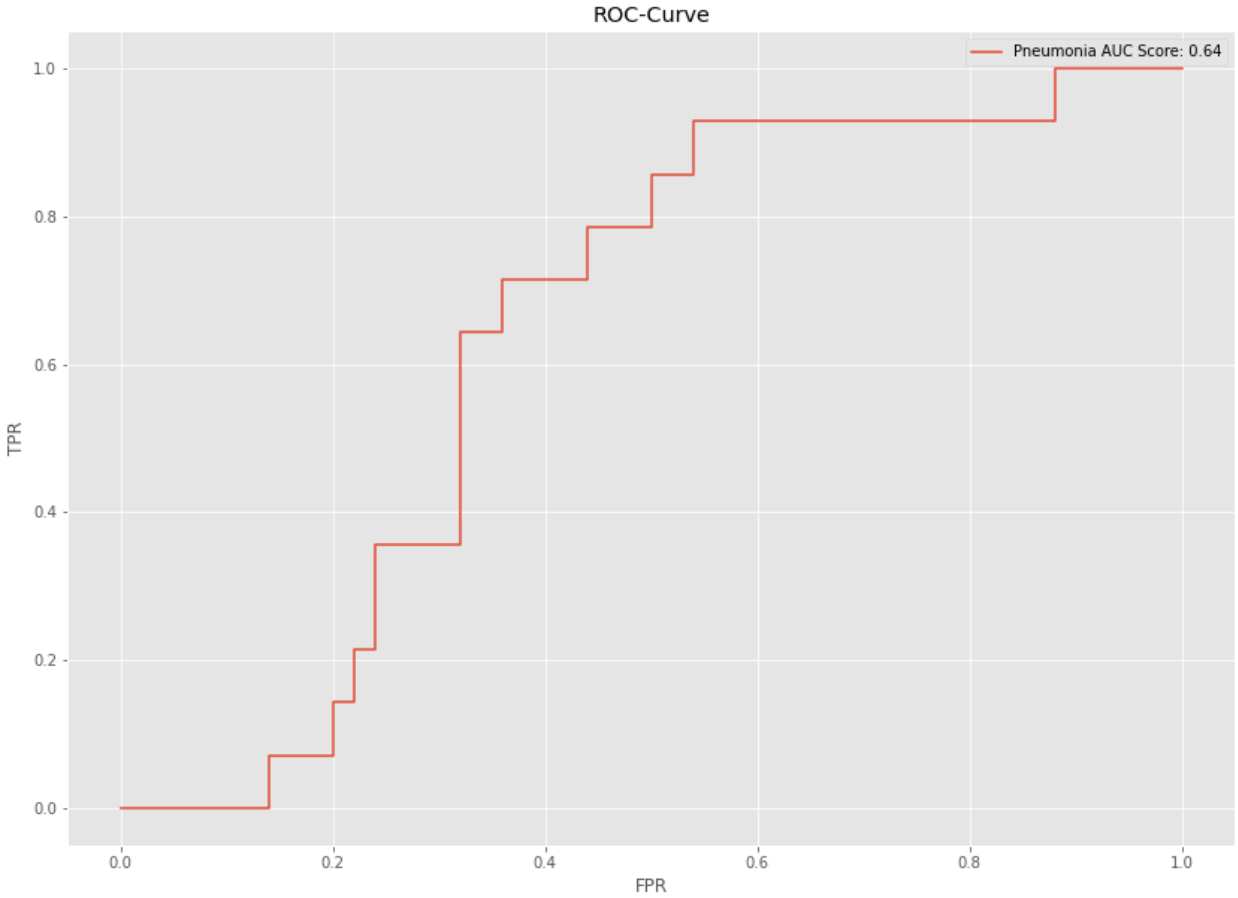
3. Algorithm Training

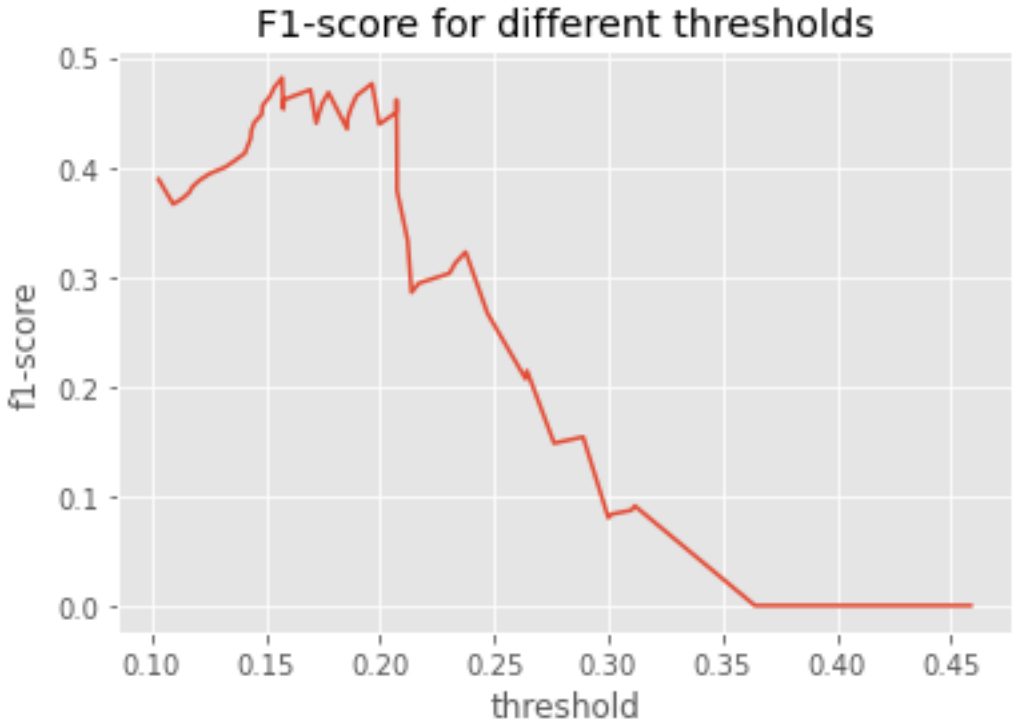
Parameters: * Types of augmentation used during training * horizontal_flip set to True, * height_shift_range set to 0.1, * width_shift_range=0.1, * rotation_range=20, * shear_range=0.1, * zoom_range=0.5

- Batch size set to 32
- Optimizer learning rate set to 3e-4
- Layers of pre-existing architecture that were frozen = 17
- Layers of pre-existing architecture that were fine-tuned: The last convolutional layer was fine-tuned
- Layers added to pre-existing architecture




PR-Curve
Pneumonia AP Score: 0.27868987752172214
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0





Final Threshold and Explanation:

The performance of the 3 different thresholds was explored using precision, recall and F1. for instance: Precision at 0.15699854493141174 threshold is 0.325. Recall at 0.15699854493141174 threshold is 0.9285714285714286. F1 score at 0.15699854493141174 threshold is 0.48148148148148157.

4. Databases

Both training and validation dataset were from NIH chest X-ray dataset. This dataset comprises of 112,120 X-ray images with disease labels from 30,805 unique patients. See the column names and object type below:

```
## Image Index                     object
## Finding Labels                  object
## Follow-up                       int64
## Patient ID                       int64
## Patient Age                      int64
## Patient Gender                  object
## View Position                   object
## OriginalImage[Width              int64
##cHeight]                          int64
## OriginalImagePixelSpacing[x y]    float64
## y]                              float64
## Unnamed: 11                     float64
## dtype: object
```

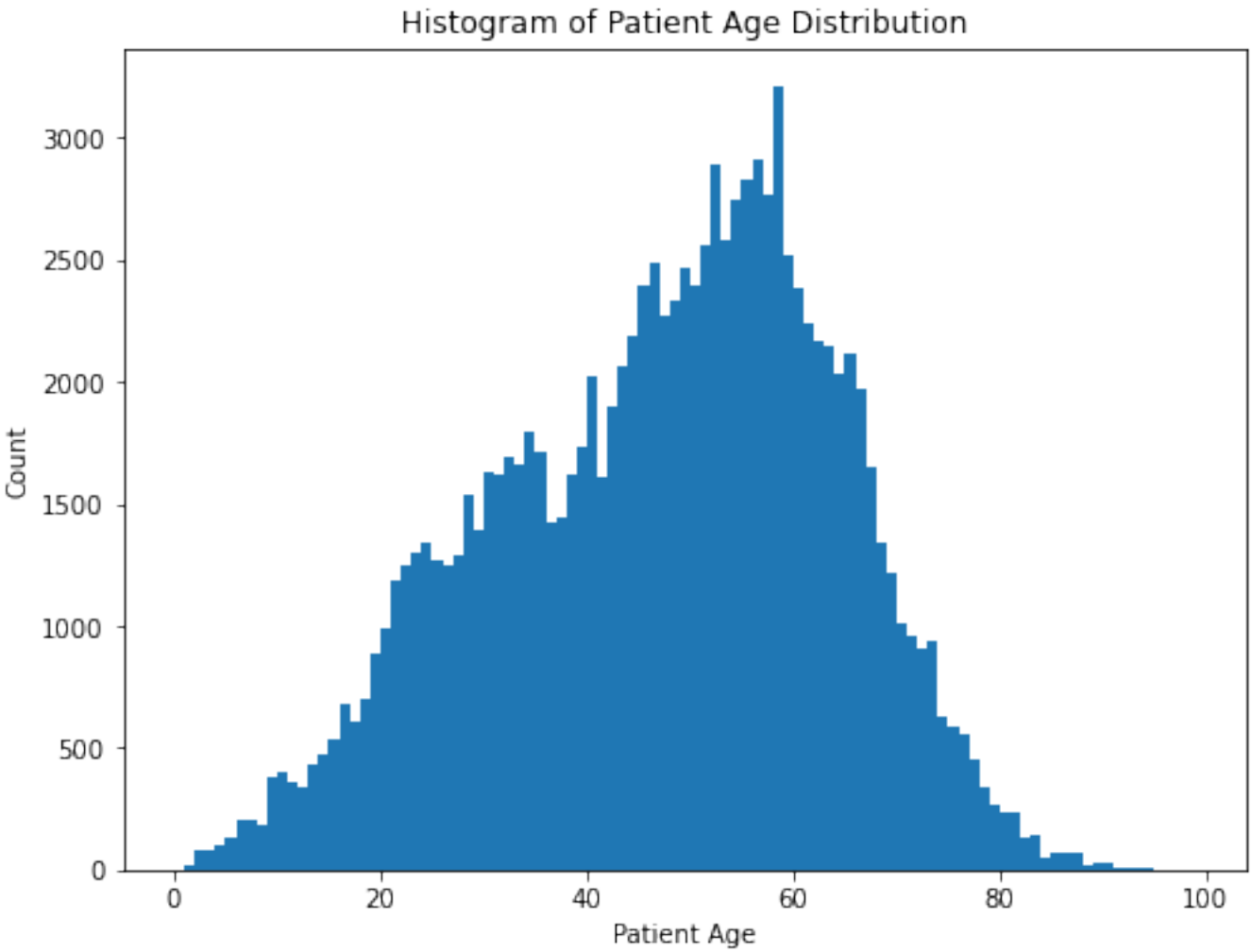


Figure 1: Age distribution of NIH chest X-ray dataset

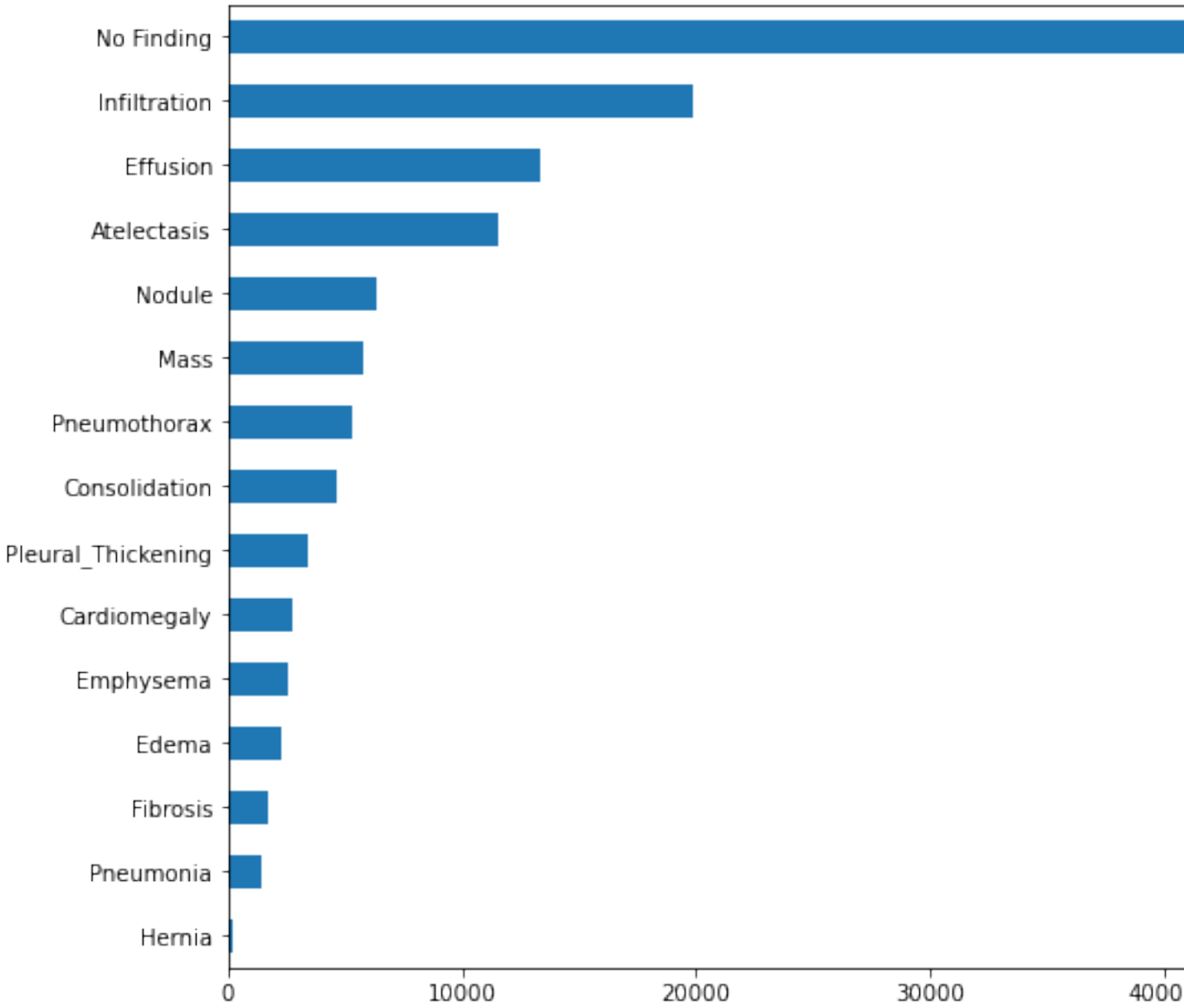


Figure 2: Occurence of unique labels

Description of Training Dataset: The training data has 2290 chest-xray images which are evenly distributed to 1145 pneumonia cases and 1145 non-pneumonia cases.

Description of Validation Dataset: The training data has 1430 chest-xray images. It contains 1144 non-pneumonia cases and 286 pneumonia cases.

5. Ground Truth

To create these ground truth, the authors had used NLP text-mine disease classifications from the associated radiological reports.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: The FDA Validation Dataset contains both female and male in the age bracket of 1 to 95 years. **Ground Truth Acquisition Methodology:** NLP text-mine disease classifications from the associated radiological reports was used to obatin ground truth.

Algorithm Performance Standard: The algorithm performance was validated using F1-score. We achieved an F1-score of 0.5 which is better than radiologist average F1-score of 0.387.