# FDA Submission

**Name:**
Savithri Brahmadathan

**Name of your Device:**
Pneumonia detection algorithm using chest x ray images

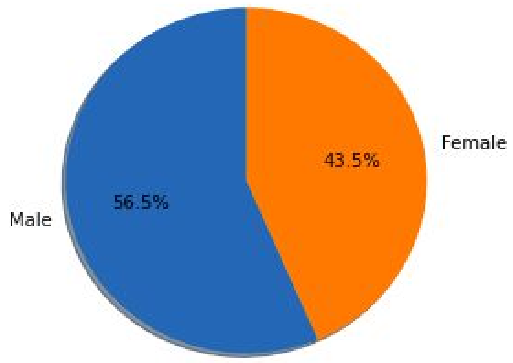
### Algorithm Description

This algorithm can be used to screen pneumonia from chest X-rays images using CNN architecture.

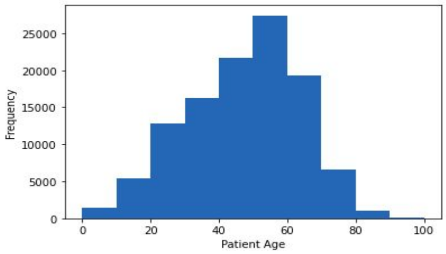
## 1. General Information

Here are some explanations regarding the dataset given:

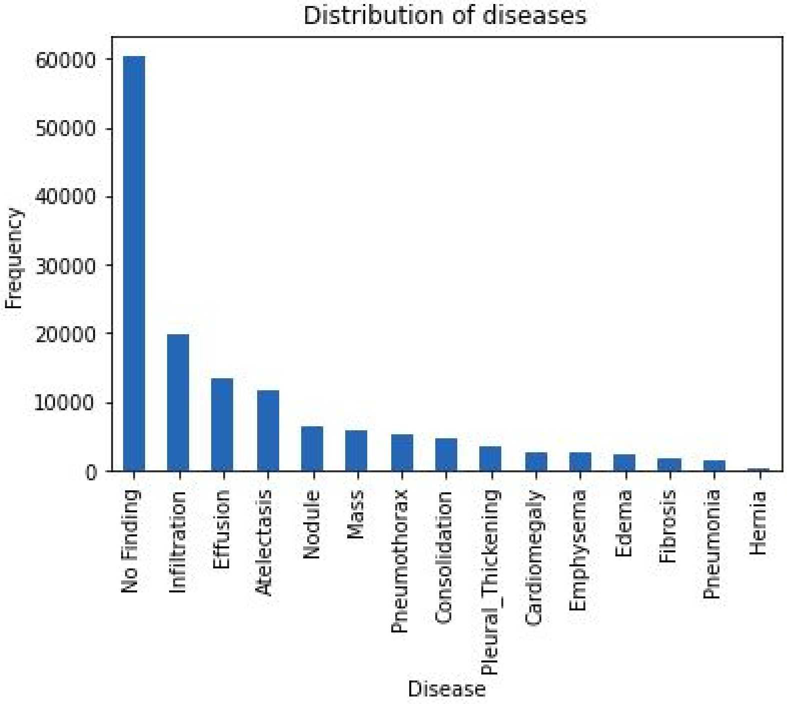
### Gender



### Age



## DIseases



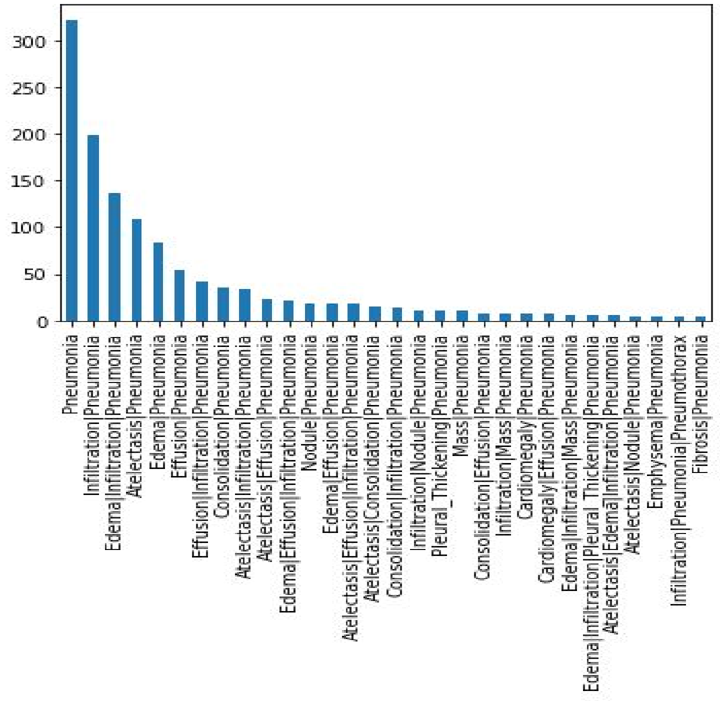
****Intended Use Statement:****

From the above diagrams,  the algorithm was trained on both male and female patients from the age range of 1-100. The patients were scanned for chest x-ray. From the data it was found that 53.8 % of patients were healthy. The rest of them were labeled with 14 other diseases including pneumonia.

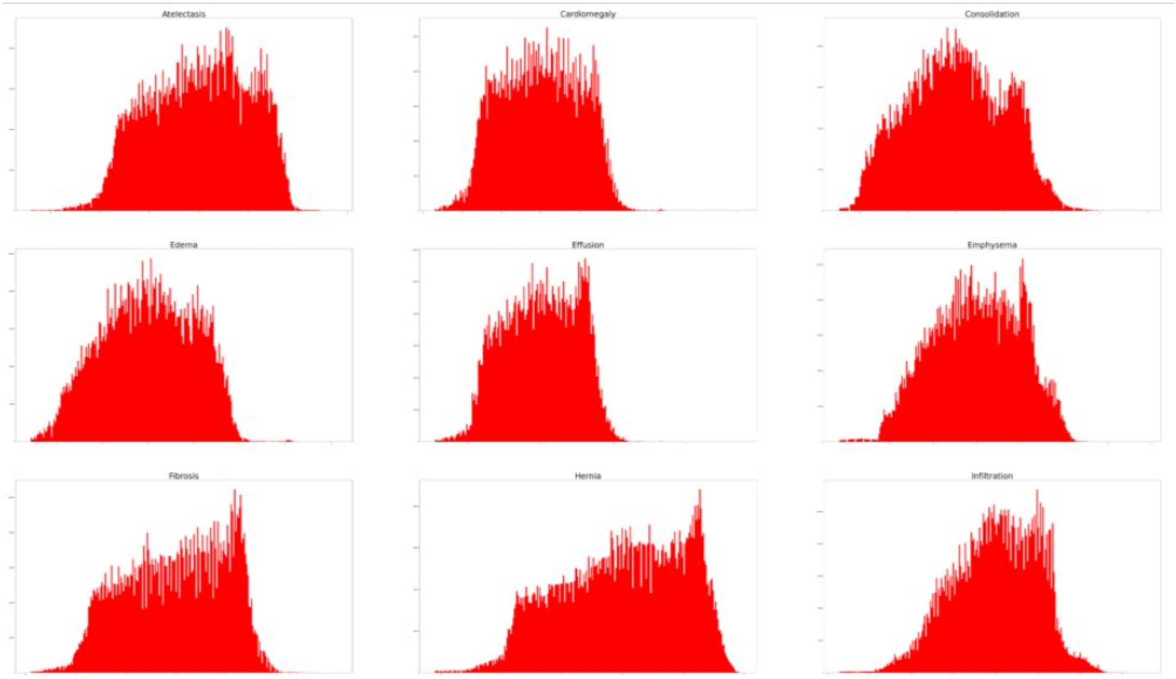
****Indications for Use:****

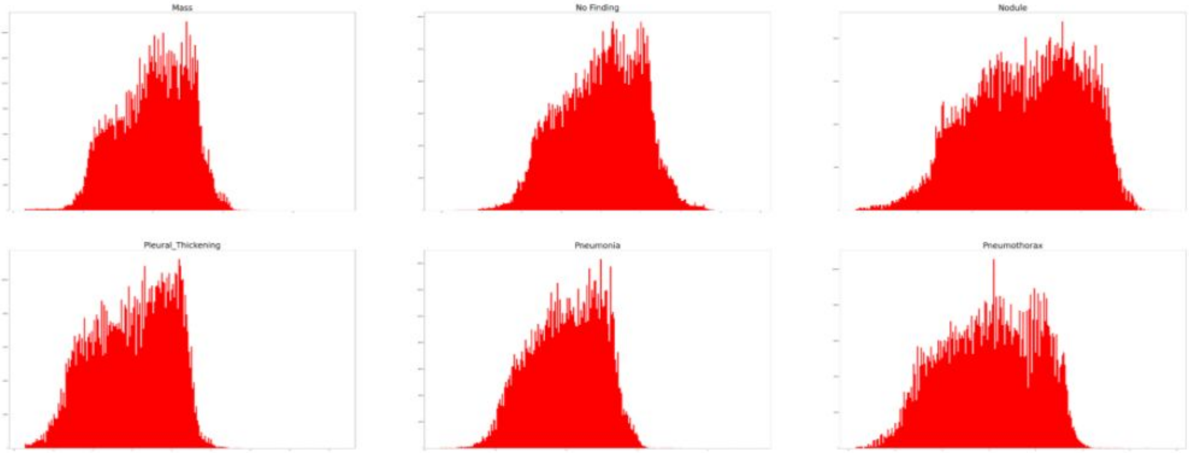
This algorithm is used to screen for pneumonia using the chest x-ray images.

**Device Limitations:**

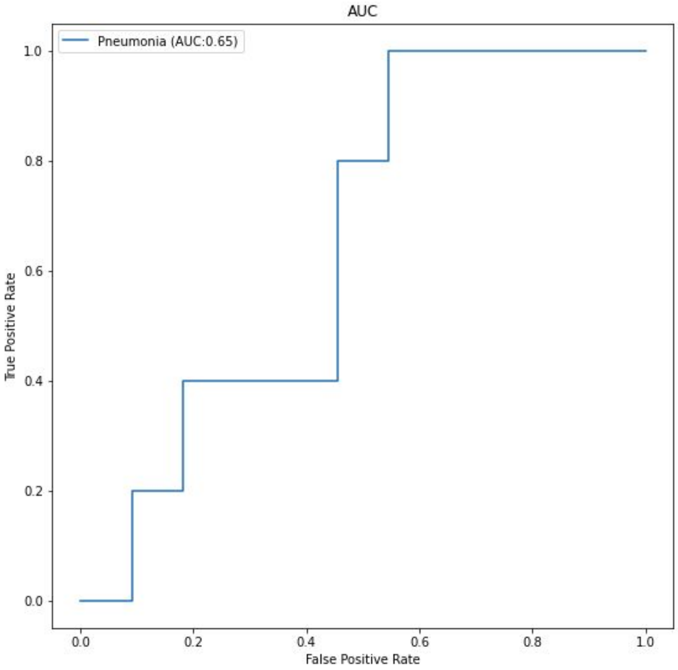


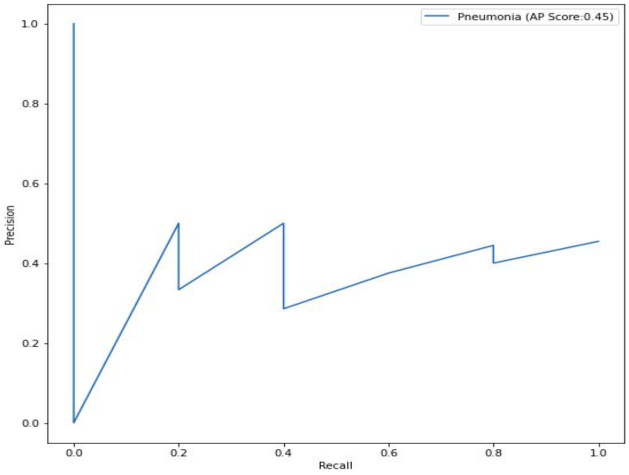
From the above diagram, we can see that there are many diseases that are comorbid with pneumonia. Several of these diseases, however, have a clearly distinct intensity profile from pneumonia. By discarding images that have a clearly different intensity profile from pneumonia, we can improve the predictive power of our model by weeding out false positives. However, there are still 3 diseases that have a similar intensity profile to pneumonia and remain as a limitation: (1) infiltration, (2) mass, and (3) pleural thickening. The presence of any of these three diseases will likely lead to the model producing a false positive.


Atelectasis
Cardiomegaly
Consolidation
Edema
Effusion
Emphysema
Fibrosis
Hernia
Infiltration


Mass
No Finding
Nodule
Pleural_Thickening
Pneumonia
Pneumothorax

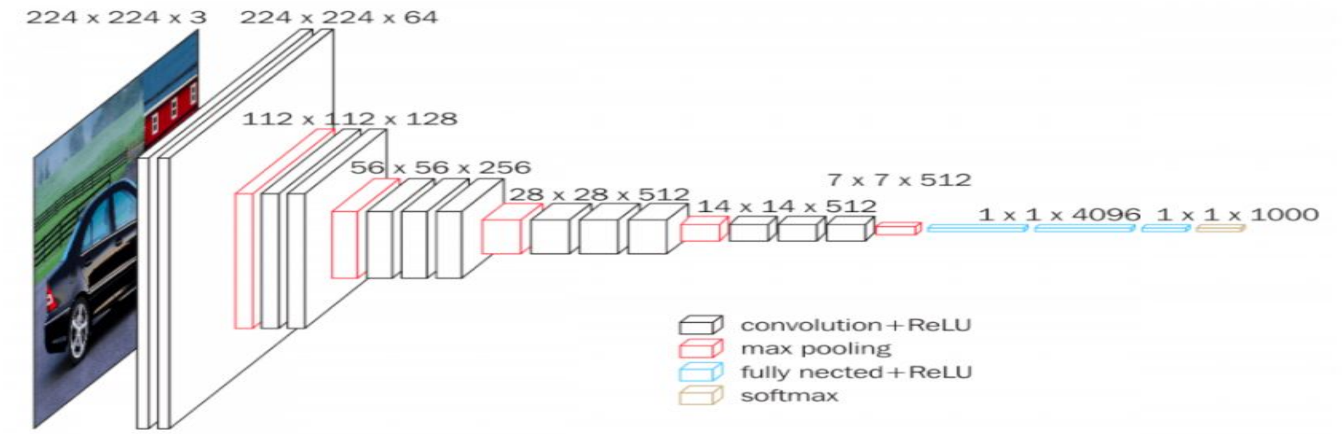
****Clinical Impact of Performance:****





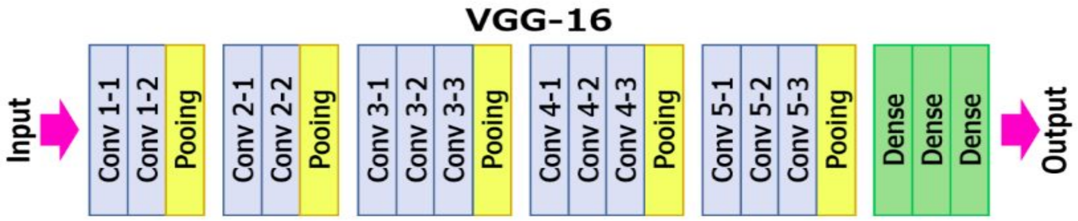
Due to the high sensitivity, this algorithm can be  used in assistance with radiologist for the screening of pneumonia from the chest x ray images.

## 2. Algorithm Design and Function

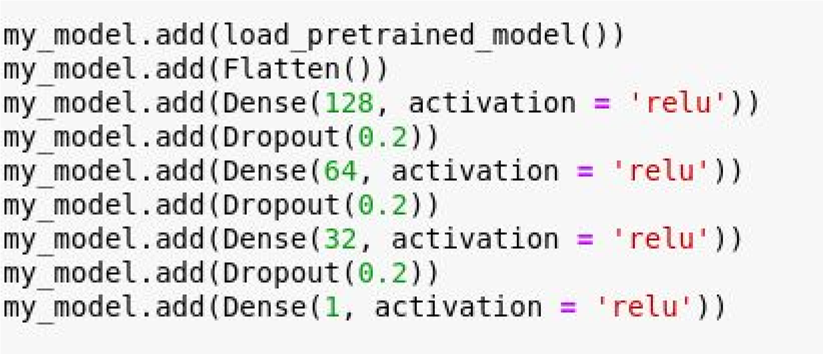


**DICOM Checking Steps:**
The chest x-ray images are in dicom files.So, we need to convert all the images into pixel size to normalize the image distribution.  We have used the Resize method to get images of size 224 pixels to resize the images. We checked whether the body part examined was chest or not if not it will print as invalid images. Similarly we also checked if modality == ‘DX’ and patient position ==  AP or PA. If not, return an invalid image. Otherwise  it will load the images.

**CNN Architecture:**



Using the pre-trained VGG-16 model as the base of the architecture,we froze all the layers except the layers from'block5_pool'. After that we added the following layers:

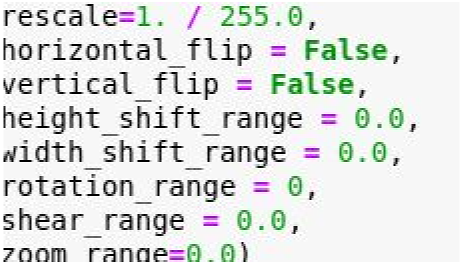


```
my_model.add(load_pretrained_model())
my_model.add(Flatten())
my_model.add(Dense(128, activation = 'relu'))
my_model.add(Dropout(0.2))
my_model.add(Dense(64, activation = 'relu'))
my_model.add(Dropout(0.2))
my_model.add(Dense(32, activation = 'relu'))
my_model.add(Dropout(0.2))
my_model.add(Dense(1, activation = 'relu'))
```

In this algorithm we added a flatten layer and four dense layers. VGG-16 is already optimized for image classification, so using this model and only changing the few output layers seemed like it would yield good results. In this algorithm we are calculating the f1 score instead finding the accuracy of the model.

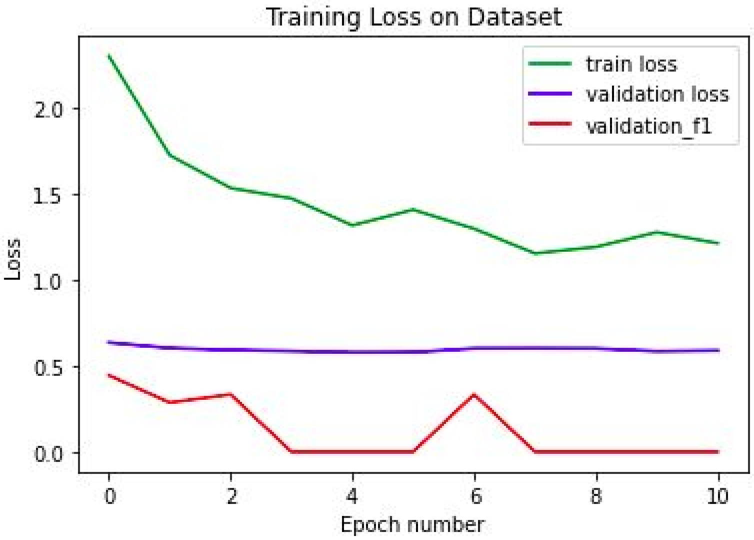
## 3. Algorithm Training

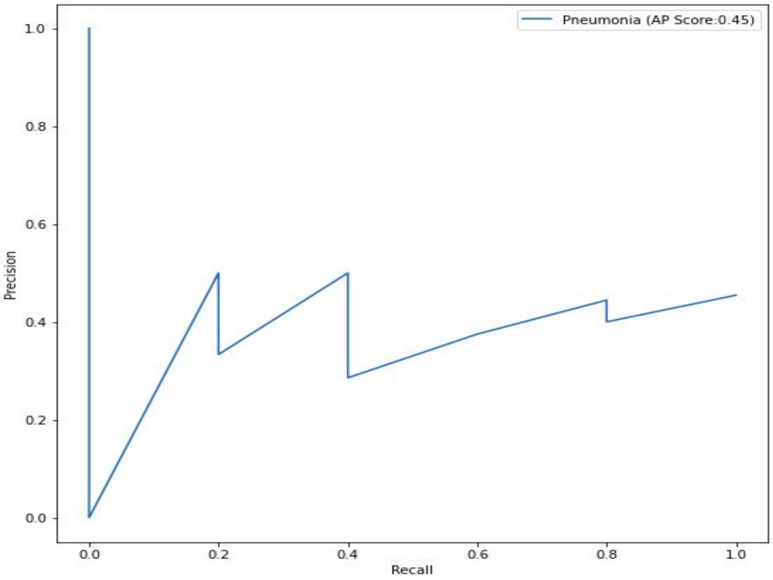
We trained the model for 100 epochs. During the training we applied horizontal_flip = True, no vertical_flip = False, height_shift_range  of 0.2, width_shift_range also 0 and a rotation_range 30 degree, shear_range of 0.2, and zoom_range of 0.2. For validation we used:



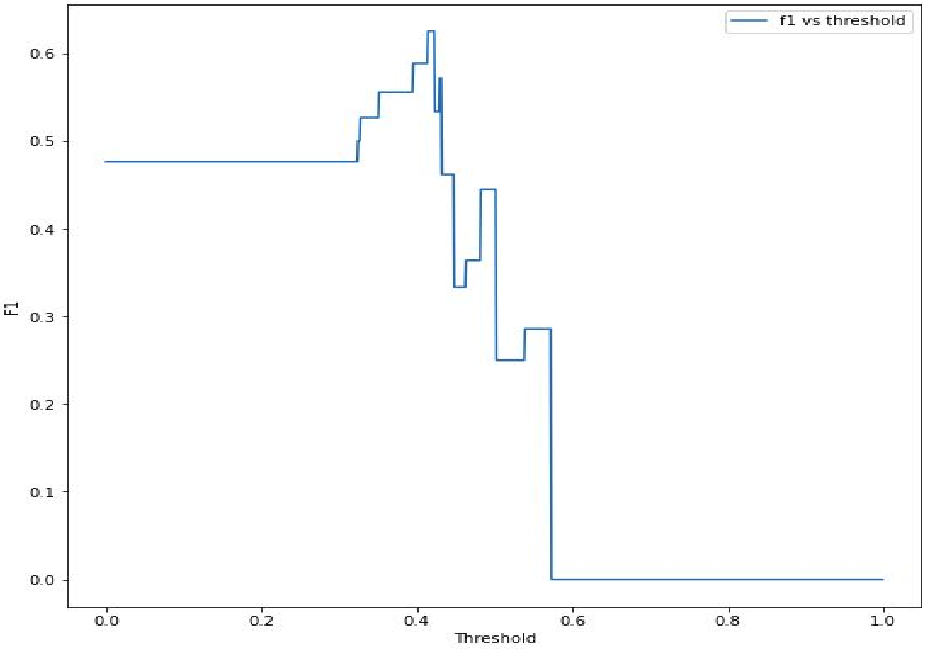
```
rescale=1. / 255.0,
horizontal_flip = False,
vertical_flip = False,
height_shift_range = 0.0,
width_shift_range = 0.0,
rotation_range = 0,
shear_range = 0.0,
zoom_range=0.0)
```

We used a batch size of 16 and adam optimizer with learning rate 1e-6  for the training.We saved the model with the best maximum val-f1 score.


Training Loss on Dataset
train loss
validation loss
validation_f1
Loss
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
Epoch number


Pneumonia (AP Score:0.45)
Precision
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Recall

**Final Threshold and Explanation:**



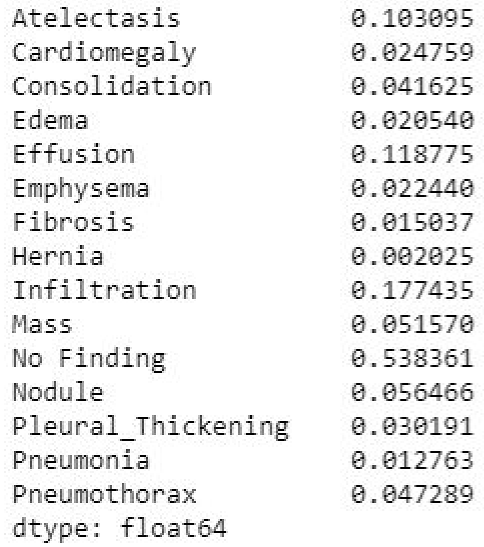
Threshold of 0.41 gives best f1 at 0.6250.

.

## 4. Databases

(For the below, include visualizations as they are useful and relevant)The dataset given from NIH contains many diseases other than pneumonia. There were 28 labels for each patient.

The 15 labels were the information regarding different  diseases.



```
Atelectasis           0.103095
Cardiomegaly          0.024759
Consolidation         0.041625
Edema                 0.020540
Effusion              0.118775
Emphysema             0.022440
Fibrosis              0.015037
Hernia                0.002025
Infiltration          0.177435
Mass                  0.051570
No Finding            0.538361
Nodule                0.056466
Pleural_Thickening    0.030191
Pneumonia             0.012763
Pneumothorax          0.047289
dtype: float64
```

**Description  Dataset:**
In the training model we split the dataset into 80% as training dataset and 20% as validation dataset. In order to get a balanced data set, we remove several negative cases so that we get 50% with pneumonia and 50% without pneumonia. We performed some augmentation in the training dataset and created a training data generator with a target size of 224x224. The validation data generator is the same as the training generator, but with no augmentation.

## 5. Ground Truth

We used NLP techniques on radiologists reports to detect whether the patient had pneumonia or not. This technique means that it is possible that some labels are incorrect due to errors in the NLP technique used. The developers of the NLP algorithm used believe an accuracy of ~90% is to be expected for this use-case.

## 6. FDA Validation Plan

**Patient Population Description for FDA Validation Dataset:**
The dataset should contain patients with both genders and in between 1-100 ages. This algorithm is not intended to be used in patients with prior history of pneumonia or any other disease such as infiltration, edema etc. These may sometimes show false positives.

**Ground Truth Acquisition Methodology:**
The images can be taken in both AN and PN position.

**Algorithm Performance Standard:**
We used the silver standard, using the weighted average of several radiologist reports. Given the results published by Rajpurkar et al. for their model CheXNet (who achieved an F1 score of 0.435), we believe that an F1 score of 0.5 would be a good minimum acceptable score as an improvement over the current state-of-the-art.