FDA Documentation

Name: Abdelrahman Rashed

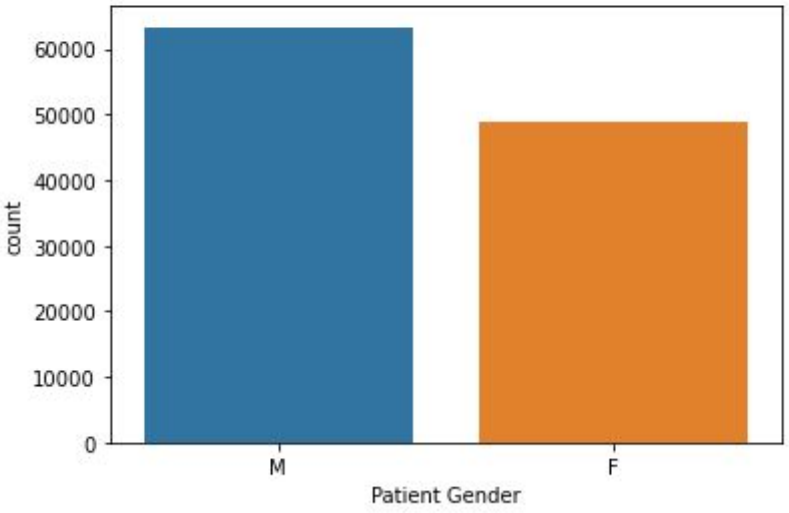
Algorithm Description: Pneumonia Detection Algorithm.

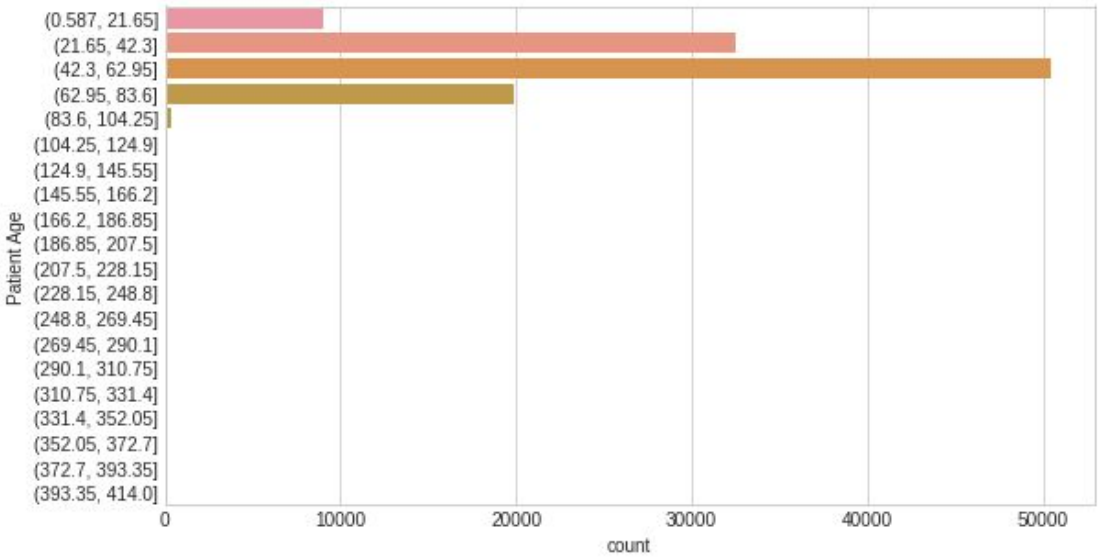
1. **General Information:**

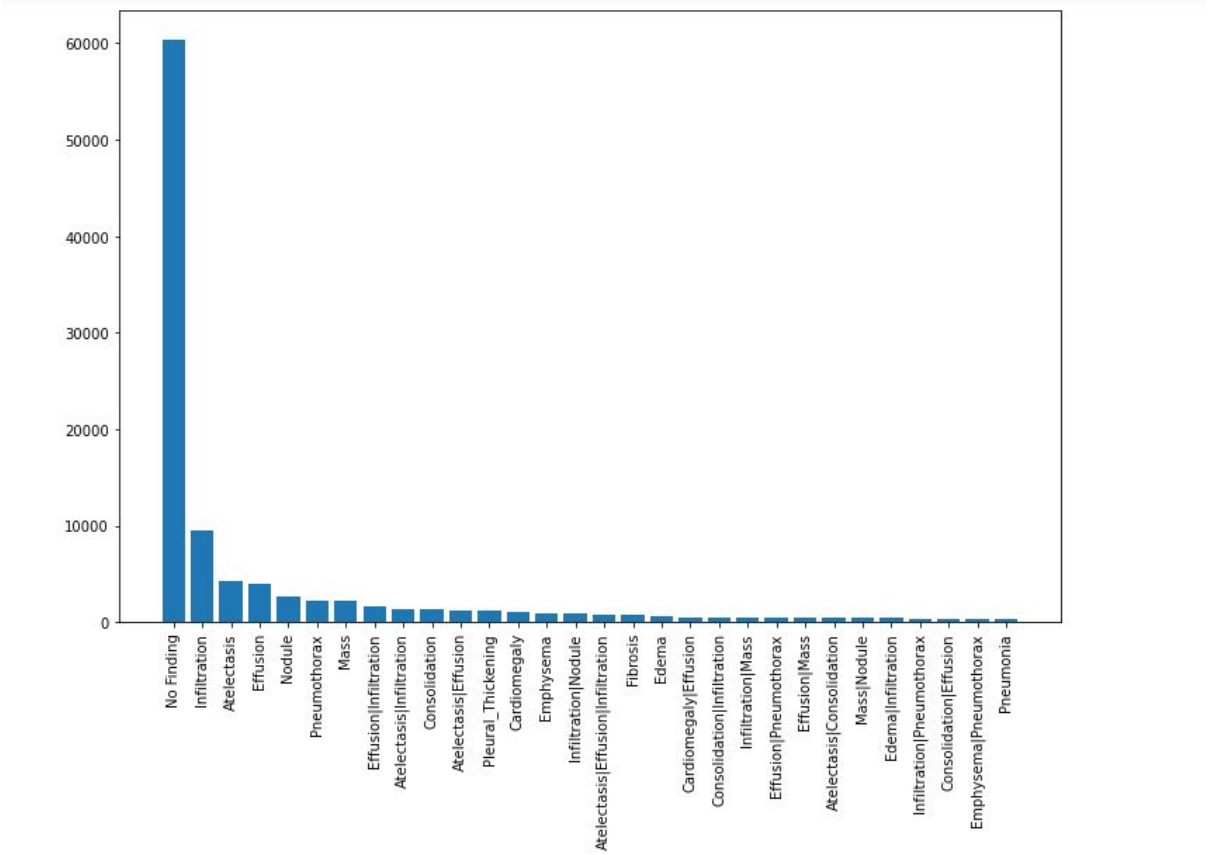
- **Intended Use Statement:** Detection of pneumonia on chest x-rays for patients of different ages.

- **Indications for Use:** This algorithm's clinical intended use is for radiologists for the detection and suspicion of pneumonia on chest x-rays. The algorithm has no limitation about the ages of the patients [0 - 100] (There are outliers in the dataset) in the dataset and it is also suitable for both gender (males & females). The algorithm depends on chest x-ray scan images as an input.

```
Number of Male in the dataset = 63340
Percentages of the Males:  56.49304316803425
Number of Females in the dataset = 48780
Percentages of the Females:  43.50695683196575
```





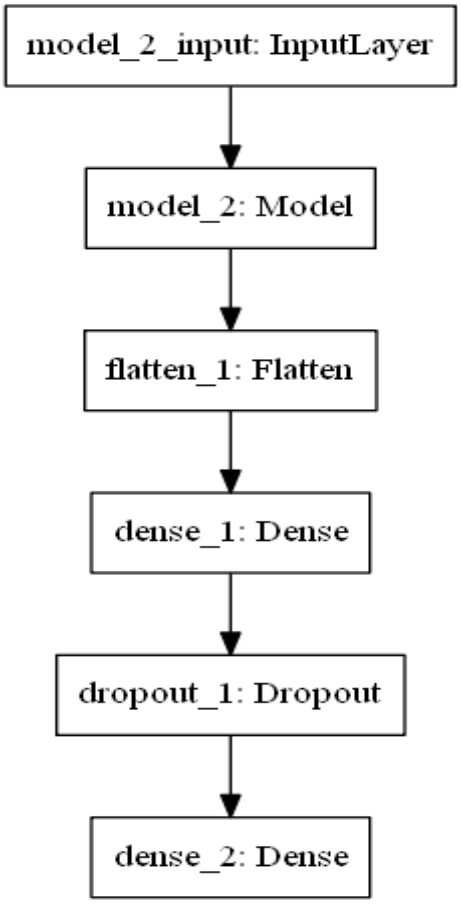


- **<u>Device Limitations:</u>** The algorithm evaluation showed that the existence of the diseases like Infiltration, Atelectasis, Effusion, Nodule, Pneumothorax, Mass, Consolidation, Pleural_Thickening, Cardiomegaly, Emphysema, Fibrosis or Edema may lead to false-positive pneumonia prediction as specificity is equal to 1. While the prediction of the Pneumonia alone is more accurate to be detected by the algorithm.

- **Clinical Impact of Performance:** The algorithm has a good result in detecting Pneumonia which can be very beneficial in automating the process of detecting Pneumonia cases directly from the X-ray chest which saves the time in the diagnosis process. The FP percentage is higher than the percentage of FN where experts will then manually double confirm the existence of Pneumonia. This might slow down the automatic process slightly, but it is better to detect False positive over the False negative.

2. Algorithm Design and Function:

- **Flowchart:**



- **DICOM Checking Steps:** The algorithm loads the images from the Dicoms files stored in the same project directory. Then, there are some checks to be made to the image once they are loaded. Checking the part of the body which the X-ray shows and it shall be 'Chest' to be eligible for the algorithm. Also checking the "Modality" and it shall be "DX". Finally for the last step, Checking the patient position and it shall be "PA" or "AP" .After that, images are extracted to be plotted. Finally I print some of the provided data like patient's age, bodypartExamined or Study description.

- **Preprocessing Steps:** First we get the numpy array output from the check_dicom function. Then I started to get the image mean and standard deviation. After that, I evaluate the processed image through subtracting the mean intensity from the actual image then dividing it by the standard deviation. After that I applied resizing to the new preprocessed image.

- **Architecture Description:** The architecture of the model I used consisted of a pretrained model (VGG16) model then I have added some layers below it for fine tuning.
    1. Flatten() layer to flatten the output of the pretrained model.
    2. Dense(1024, activation = 'relu') layer with activation function "ReLU" to combine the recognized features by the pretrained model.
    3. Dropout(0.2) layer to avoid overfitting.
    4. Dense(1, activation = 'sigmoid') layer for the final binary classification (pneumonia or Not pneumonia)

3. Algorithm Training:

- **Types of augmentation used during training:**

```
my_idg = ImageDataGenerator(rescale = 1./255,
                            width_shift_range = 0.1,
                            height_shift_range = 0.1,
                            shear_range= 0.1,
                            zoom_range = 0.1,
                            horizontal_flip = True,
                           vertical_flip  = False,
                           rotation_range = 20)

return my_idg
```

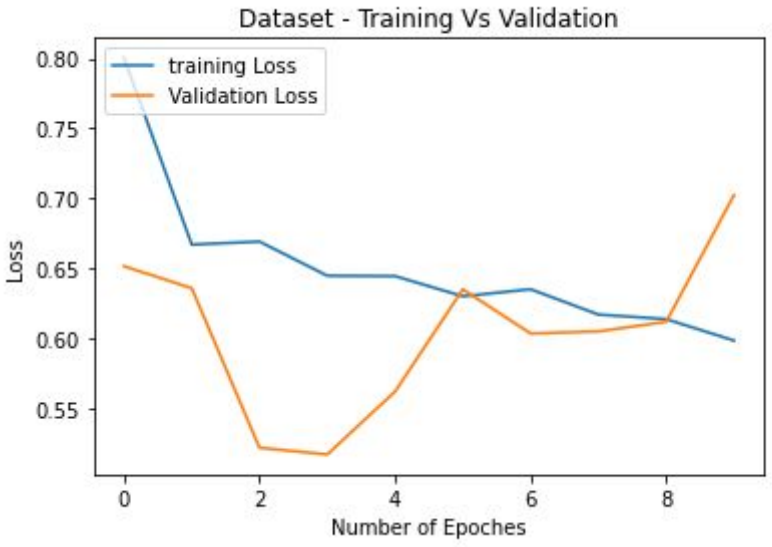
- **Batch size :** (64)

- **Optimizer learning rate:** Adam(lr=0.0001, decay=1e-5)

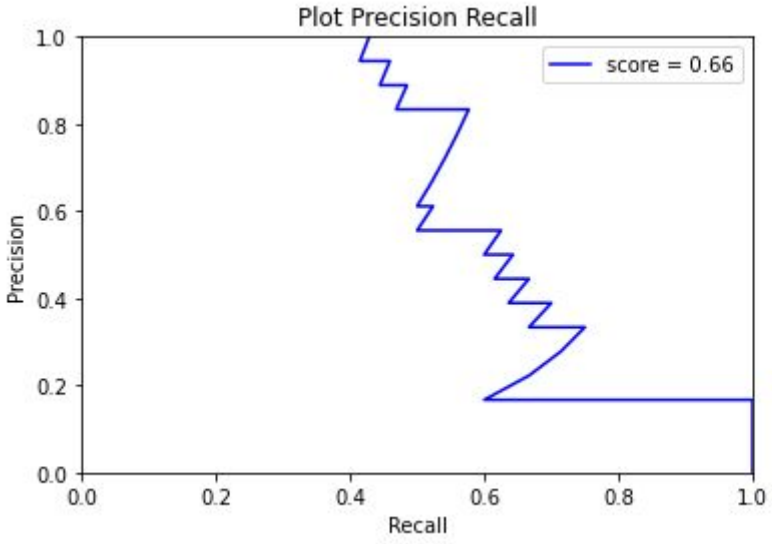
- **Layers of pre-existing architecture that were frozen:** I have designed the algorithm using the pretrained model (VGG16) model. Then after the pre trained model I have added the following layers to be fine-tuned. The model is then trained for 10 epochs.

```
# Building the model with transfer learning and added new layer for classification
my_model = Sequential()

vgg16_model = load_pretrained_model()

# Addig from the above the VGG16 model
my_model.add(VGG16_Conv_model)

# FC layer
# Flatten the VGG16 model output  -- shape of last block = (7,7,512)
my_model.add(Flatten())

# Add the fully-connected layer (Dense layer) - To combine the recongized feature by VGG16
my_model.add(Dense(1024, activation = 'relu'))

# Add the droup-out layer to avoid overfitting
my_model.add(Dropout(0.2))

# Add a final sigmoid layer for classification
my_model.add(Dense(1, activation= 'sigmoid'))

return my_model
```

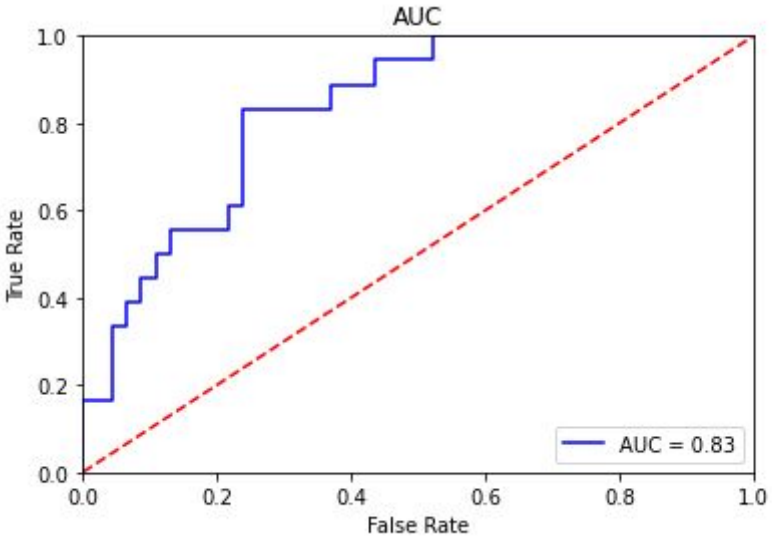
- **Algorithm training performance:**



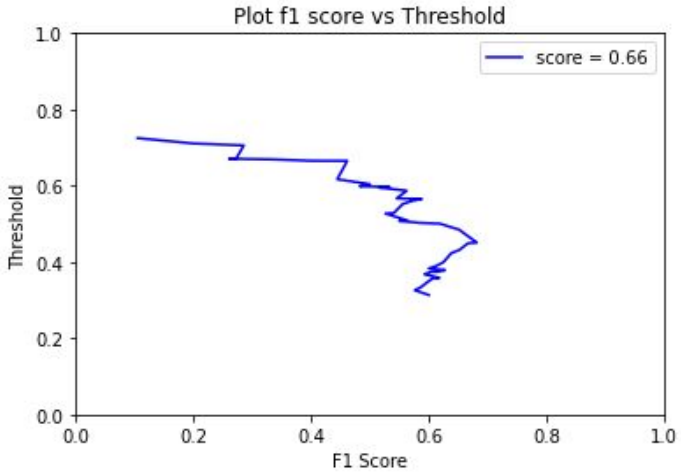
- **P-R curve:**



- **Final Threshold:** (Final threshold = 0.6)
- **Final Accuracy:** (Final Accuracy = 0.78125)
- **Accuracy Curve:**



- **F1-Threshold Curve:**



4. Database

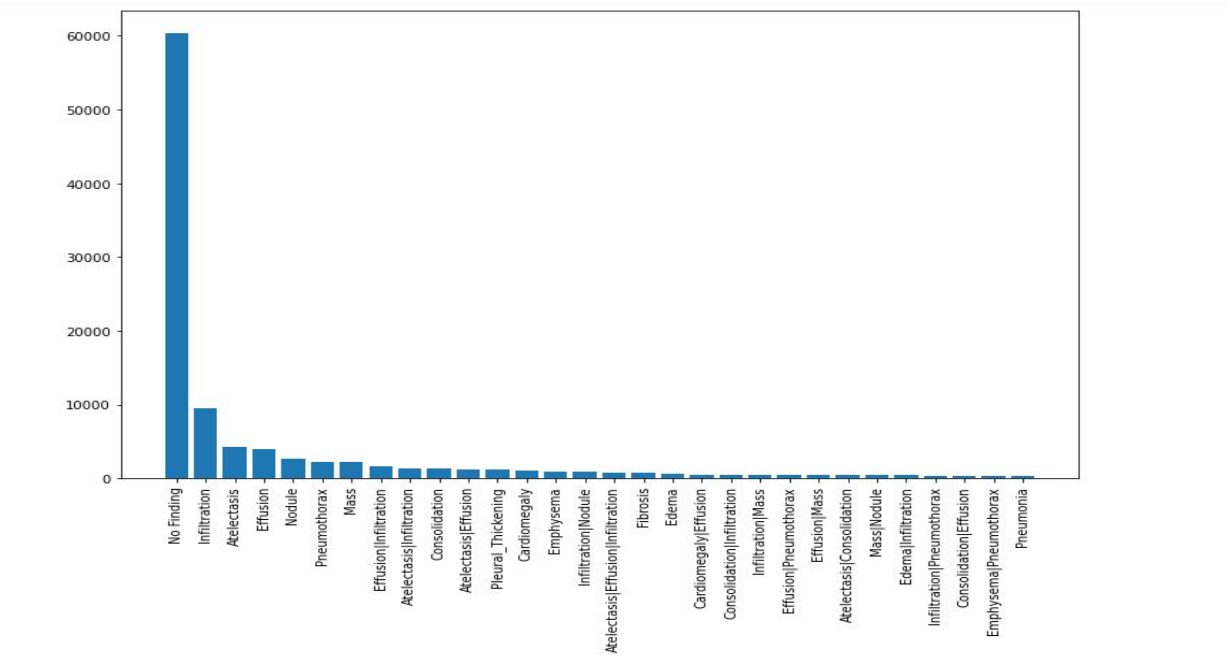
- The NIH dataset contains 112,120 images with disease labels from 30,805 unique patients. The labels are more than 90% accurate and valid for supervised learning. The dataset's labels differentiate between 14 diseases and the 'No Finding' label.

- The names of disease are:
    - Atelectasis
    - Consolidation
    - Infiltration
    - Pneumothorax
    - Edema
    - Emphysema
    - Fibrosis
    - Effusion
    - Pneumonia
    - Pleural_thickening
    - Cardiomegaly
    - Nodule Mass
    - Hernia
    -

- **Training Dataset:** The below images describe the distribution of the diseases in the dataset, Patient Gender Distribution, patients ages distribution and ages of patients suffer from Pneumonia. And Regarding the training set itself my model. I have splitted the dataframe into training dataframe and validation data frame. Where the percentage for this splitting was 80% for train data and 20% for validation data. I have also modified the training dataset to be a balanced dataset where the number of patients who have "Pneumonia" equals the number of patients who have "No Pneumonia".

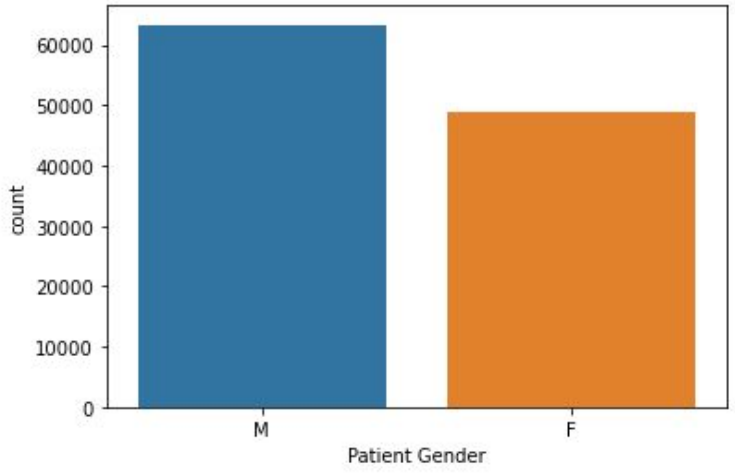
- **<u>Validation Dataset:</u>** For the validation dataset I have created it to be an unbalanced dataset where the number of the patients suffering from “Pneumonia” is the number of the patients that have “No Pneumonia”.

- **<u>NIH dataset benefits and limitations:</u>**
- Below I have plotted some distribution of the data inside the dataset.
- **<u>Dataset diseases distribution:</u>**

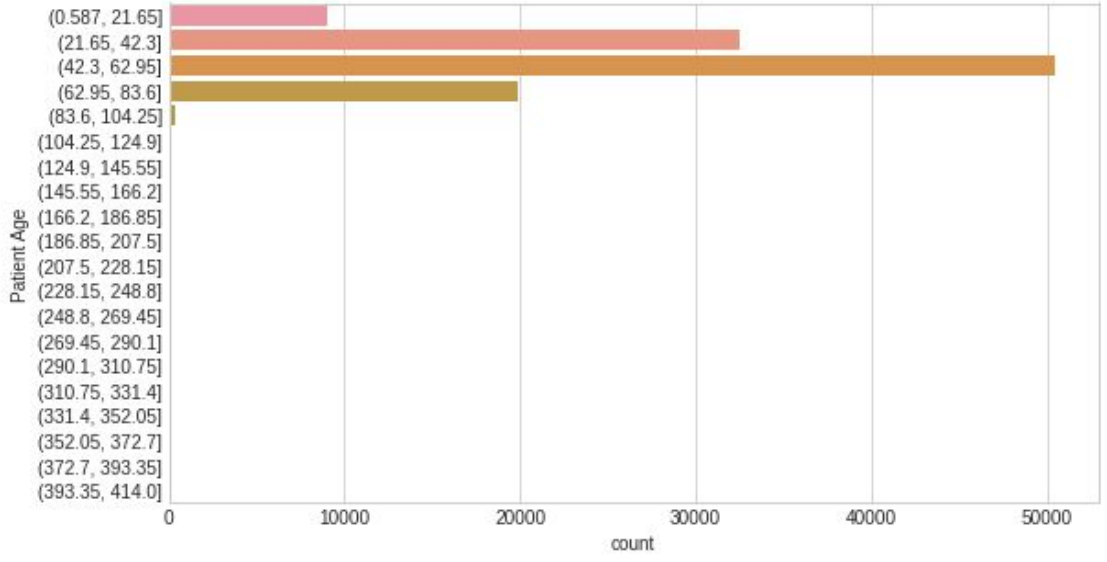


Patient Gender Distribution:

```
Number of Male in the dataset = 63340
Percentages of the Males:  56.49304316803425
Number of Females in the dataset = 48780
Percentages of the Females:  43.50695683196575
```

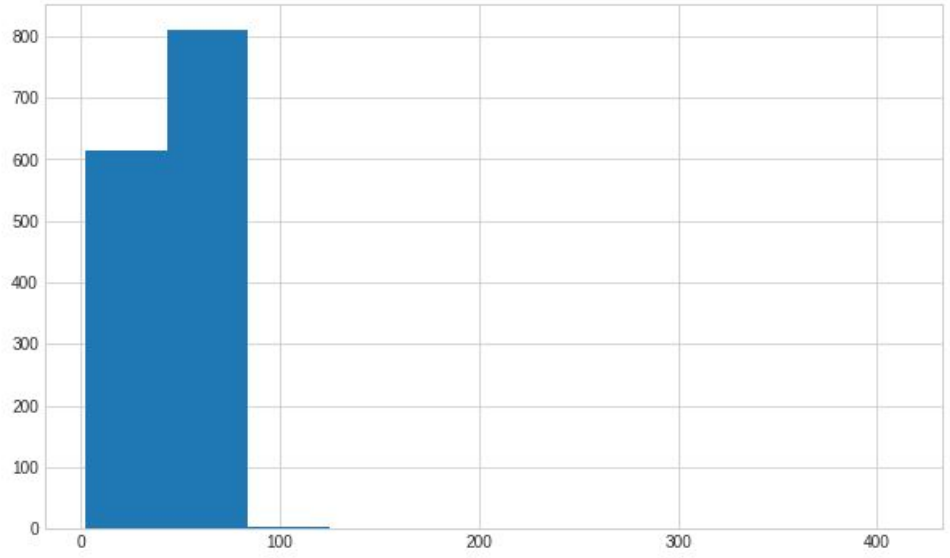


- **Patients Ages Distribution:**



- **Ages of patients suffer from Pneumonia**

```
Out[14]: (array([615., 811.,   4.,   0.,   0.,   0.,   0.,   0.,   0.,   1.]),
          array([  2.,  43.,  84., 125., 166., 207., 248., 289., 330., 371., 412.]),
          <a list of 10 Patch objects>)
```



- Distribution of the age of the patient who suffer from pneumonia

5. Ground Truth:

- The Algorithm is not 100% accurate for detecting the positive Pneumonia cases. It might predict false positive cases of pneumonia. Therefore, the positive cases shall be re-checked by the experts to double confirm them. However, the algorithm is more accurate in detecting 'No Finding' cases.

- The biggest limitation of the NIH dataset is that the image labels were NLP-extracted so there could be some erroneous labels but the benefits of the NLP labeling accuracy is estimated to be more than 90%.

- Chest x-ray radiology reports are not anticipated to be publicly shared.

- There are a very limited number of disease region bounding boxes.

6. FDA Validation Plan

- **Patient Population Description for FDA Validation Dataset:** The dataset for the validation of the algorithm shall have chest x-ray for patient's age range from 0 to 100 years and for the both gender (male & female). Also the dataset shall guarantee that it does not have a patient with prior 'Pneumonia' history and patients that suffer from other diseases like Infiltration, Atelectasis, Effusion, Nodule, Pneumothorax, Mass, Consolidation, Pleural_Thickening, Cardiomegaly, Emphysema, Fibrosis or Edema shall be examined again with experts (Radiologists) for false positive classification.

- **Ground Truth Acquisition Methodology:** Applying the intended use for this algorithm where it is used as a tool to help the radiologist to diagnose Pneumonia.
- For ground truth, the silver standard approach of using several radiologist wool is more optimal for this algorithm. This is because is really hard for radiologist to identity quickly the existence of 'Pneumonia' or not from the chest X-rays

- **Algorithm Performance Standard:** I chose (F1 score) metric to measure the performance of the model based on the proved literature in this Link. And the value of the F1 score to use is (0.435).