

# FDA Submission

**Your Name:** Mercedes Riveira

**Name of your Device:** Pneumonia Detector

## Algorithm Description

### 1. General Information

**Intended Use Statement:** Assisting radiologist with classifying pneumonia or no-pneumonia images within a screening program.

**Indications for Use:** This algorithm is intended for use in chest X-Ray images, taken at both positions PA and AP in people aging 80 or less years. It works better. It is not concluded if its performance is worsened in patients with other diseases present.

**Device Limitations:** The algorithm was not trained for patients over 80 years old. The images must be chest X-Rays in PA or AP views. It has been seen that when fibrosis and pleural thickening diseases are comorbid with pneumonia, the algorithm's sensitivity decreases, so it is not recommended to use it for patients with these other diseases.

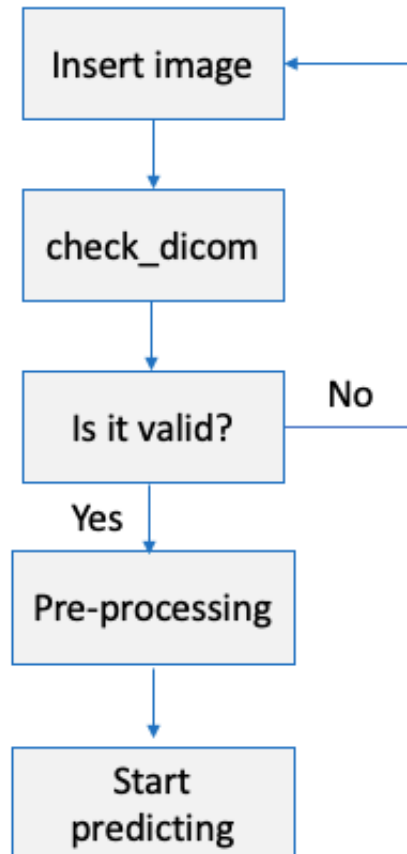
### Clinical Impact of Performance:

- Threshold: 0.4
  - Accuracy: 85%
  - Precision: 17%
  - Recall: 93% (sensitivity)

The threshold has been weighed in favour of recall (sensitivity), which means that most of the true positive cases (TPR = Sensitivity) are being well classified at the expense of classifying much of the negative cases as positive cases (that is, a high False Positive Rate and therefore a low specificity)

Since  $\text{recall} = \text{TP} / (\text{TP} + \text{FN})$ , if recall is closer to 1, FN is approximately zero (no false negatives). This means that when the algorithm runs an image and returns a **negative** result, you can be pretty sure that the result is truly a negative. This is so useful in **screening** situations (very important nowadays with COVID-19).

### 2. Algorithm Design and Function



**DICOM Checking Steps:** When an image is fed into the algorithm, the `check_dicom` function reads the DICOM file and checks the important fields for our device (Body part examined, modality and patient position). If the image is from the CHEST, in modality DX and taken in a PA or AP position, then the DICOM image (as a pixel array) is returned.

**Preprocessing Steps:** this function takes the image (pixel array) obtained from the `check_dicom` function, its mean, standard deviation and size (image size corresponding with that of VGG16 input image size of (1,224,224, 3)) as arguments, and returns the same image but processed. This processing consists of a rescaling and resizing. Finally, the valid images that are preprocessed are fed as an input to the algorithm.

**CNN Architecture:** This algorithm is built over the first 17 layers of a pretrained model (VGG16 architecture):

```

Downloading data from https://github.com/fchollet/deep-learning-models/releases/download/v0.1/vgg16_weights_tf_dim_ordering_tf_kernels.h5
553467904/553467096 [=====] - 9s 0us/step
input_1 False
block1_conv1 False
block1_conv2 False
block1_pool False
block2_conv1 False
block2_conv2 False
block2_pool False
block3_conv1 False
block3_conv2 False
block3_conv3 False
block3_pool False
block4_conv1 False
block4_conv2 False
block4_conv3 False
block4_pool False
block5_conv1 False
block5_conv2 False
block5_conv3 True
block5_pool True
Model: "vgg16"

```

Layer (type)	Output Shape	Param #
input_1 (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: 126,002,664
Non-trainable params: 12,354,880

```

Then, the following layers were added: flattening layer, three dropouts (at 0.5) to prevent overfitting, and three dense layers (of 512, 256 and 1 parameters, both firsts relu, and the last one a sigmoid).

Model: "sequential\_1"

Layer (type)	Output Shape	Param #
====		
model_1 (Model)	(None, 7, 7, 512)	1471468
8		
flatten_1 (Flatten)	(None, 25088)	0
dropout_1 (Dropout)	(None, 25088)	0
dense_1 (Dense)	(None, 512)	1284556
8		
dropout_2 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 256)	131328
dropout_3 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 1)	257
====		
Total params: 27,691,841		
Trainable params: 15,336,961		
Non-trainable params: 12,354,880		

### 3. Algorithm Training

Parameters:

- Types of augmentation used during training:

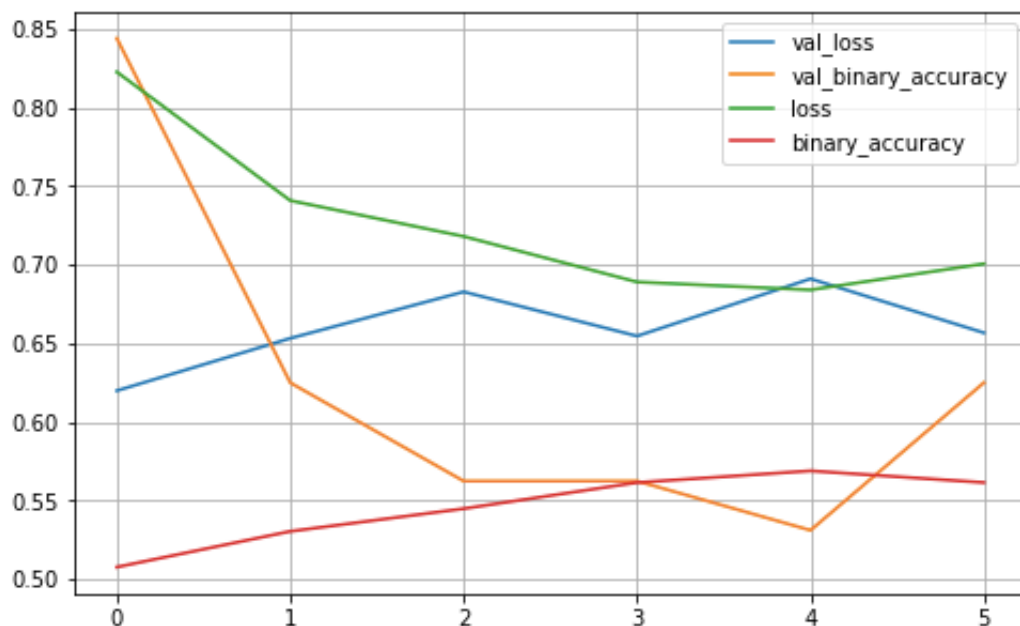
```
# Train:
train_idg = ImageDataGenerator(rescale=1. / 255.0,
                                horizontal_flip = True,
                                vertical_flip = False,
                                height_shift_range = 0.1,
                                width_shift_range = 0.1,
                                rotation_range = 20 ,
                                shear_range = 0.1,
                                zoom_range= 0.1)
```

- Batch size:
  - Train: 90
  - Validation: 32
- Optimizer learning rate:
  - Adam(lr = 1e-4)
- Layers of pre-existing architecture that were frozen: first 17
- Layers of pre-existing architecture that were fine-tuned:

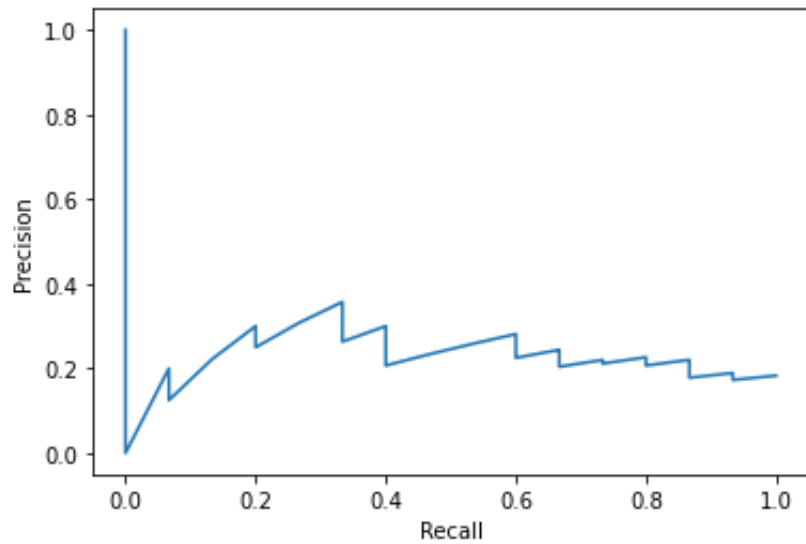
```
block5_conv3 True
block5_pool True
```

- Layers added to pre-existing architecture: flattening layer, three dropouts (at 0.5) to prevent overfitting, and three dense layers (of 512, 256 and 1 parameters, both firsts relu, and the last one a sigmoid).

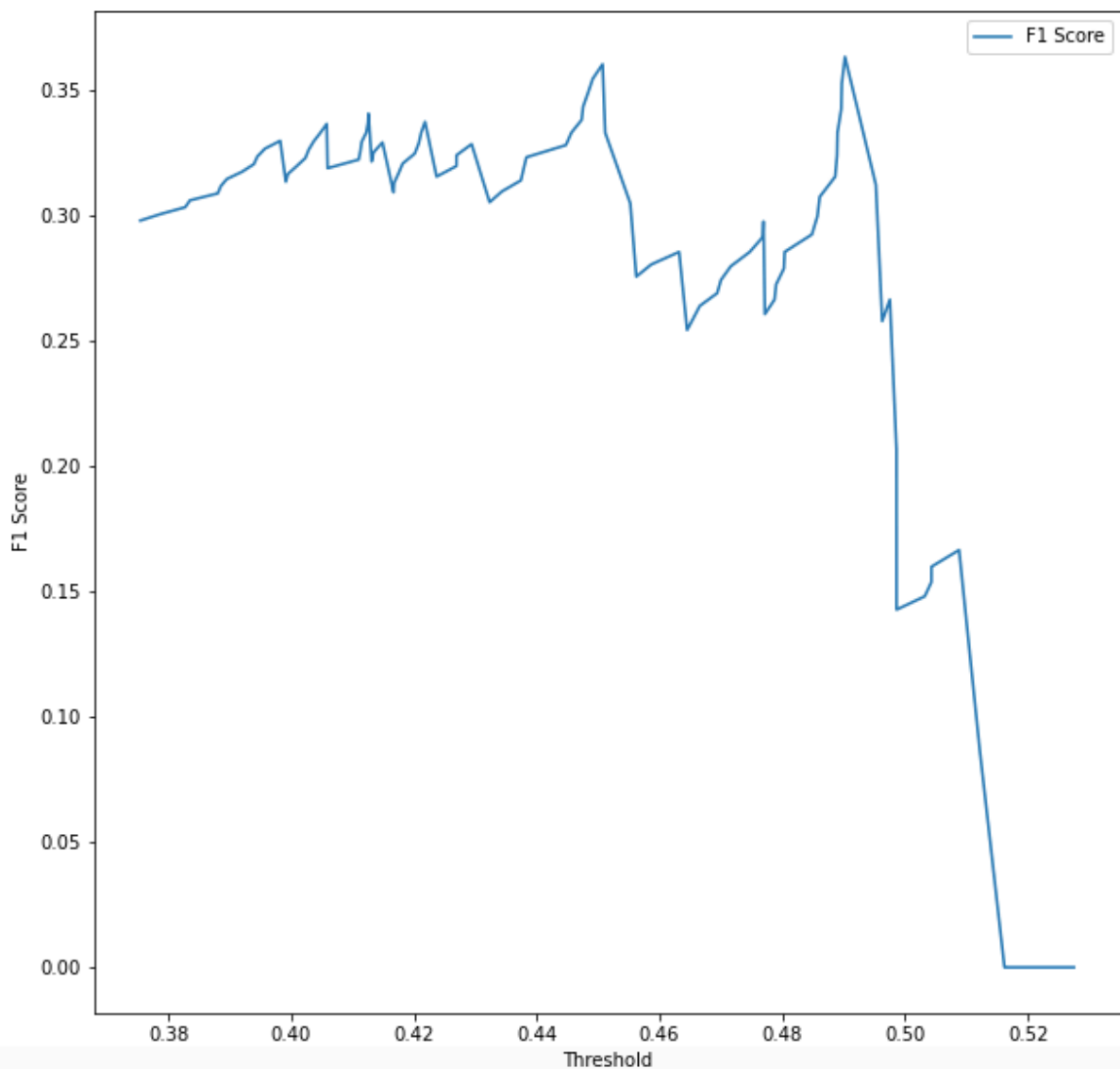
## Algorithm training performance visualization



**Insert P-R curve**



**Final Threshold and Explanation:** The final threshold was set at 0.4 because this number favours recall. Since this is a medical algorithm, it is very important to classificate all true positives at expenses of classifying some negatives as positives (false positives). The highest F1 score is 0.36 and corresponds to a threshold of 0.44.



## 4. Databases

The dataset was split following the 80–20% rule, that is: 80% for training and 20% for validation.

**Description of Training Dataset:** The proportion of Pneumonia cases in the training set is 1.2% (1145 cases out of 89696 images). This set was reshaped in order to balance the proportion of pneumonia and no pneumonia cases (50%–50%). Finally, our training dataset contains 2290 images, 1145 showing pneumonia.

**Description of Validation Dataset:** The proportion of Pneumonia cases in the validation set is 1.2% (286 cases out of 22424 images), the same as the training set. This set was reshaped in order to adjust the proportion of pneumonia and no pneumonia cases to 20%–80%.

## 5. Ground Truth

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.

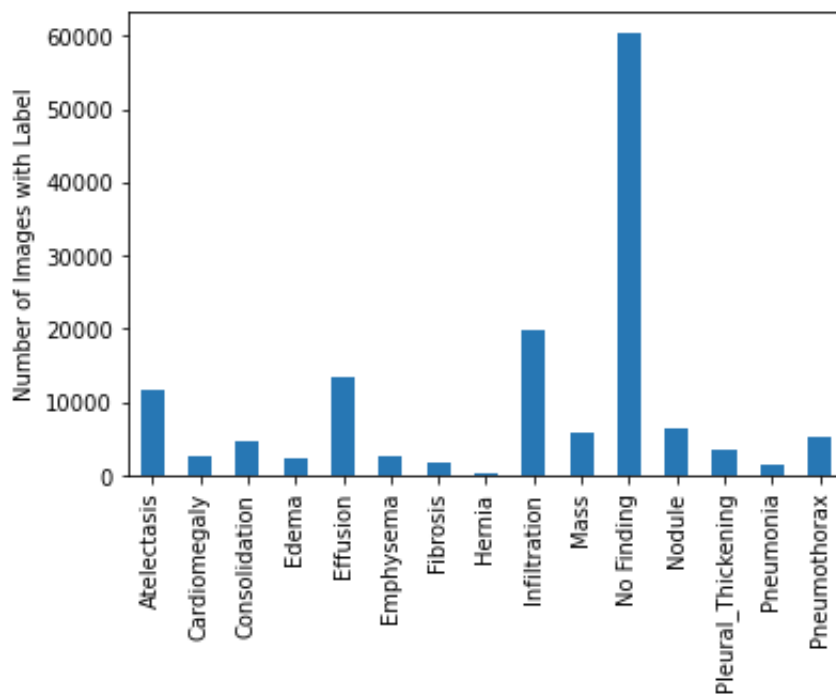
The biggest limitation of this dataset is that image labels were NLP–extracted so there could be some erroneous labels but the NLP labeling accuracy is estimated to be >90%.

## 6. FDA Validation Plan

### Patient Population Description for FDA Validation Dataset:

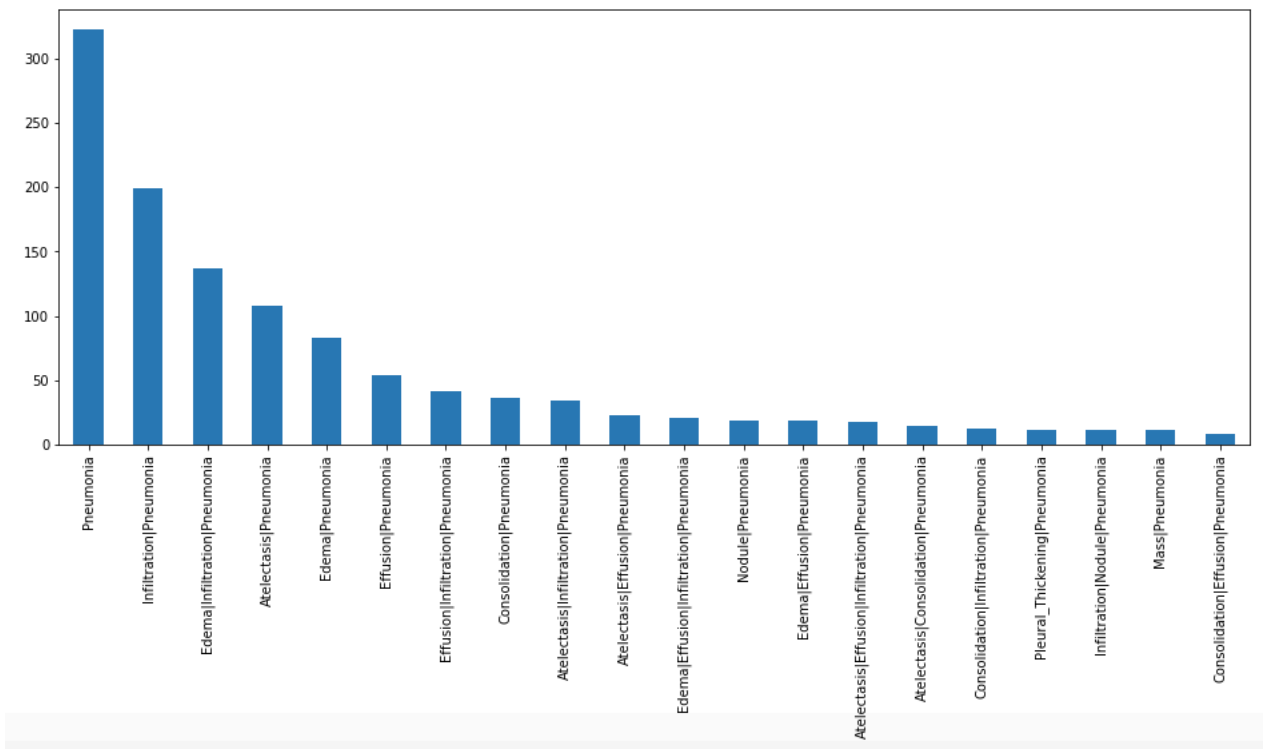
This dataset contains 112,120 x-ray chest images, obtained from the NIH database. The distribution of the diseases (figure below) shows that the 'No-finding' label is the most common (53%) followed by Infiltration, Effusion and Atelectasis. Pneumonia only constitutes 1.2% of the whole dataset

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

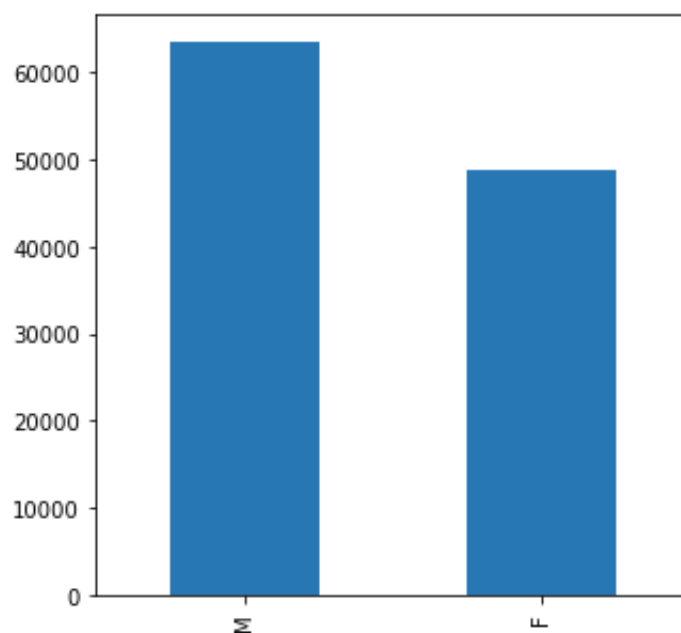


As well, the 30 most common concurrences are shown in the figure below. Stands our infiltration, followed by edema + infiltration and so on.

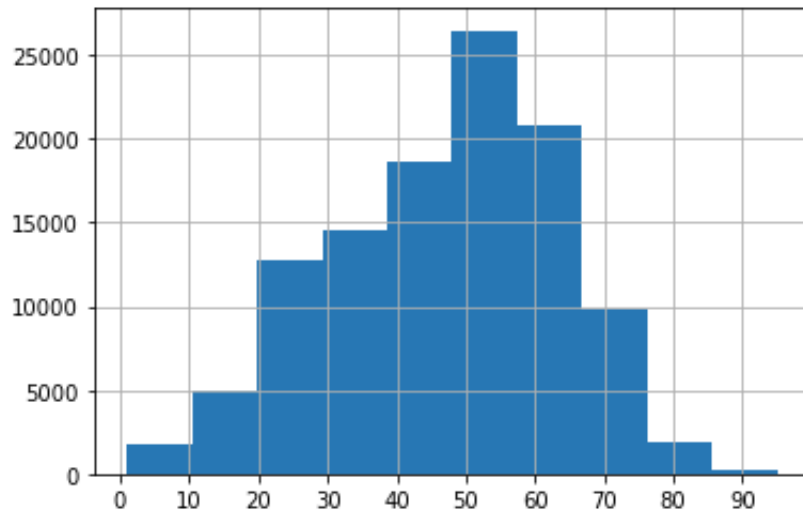




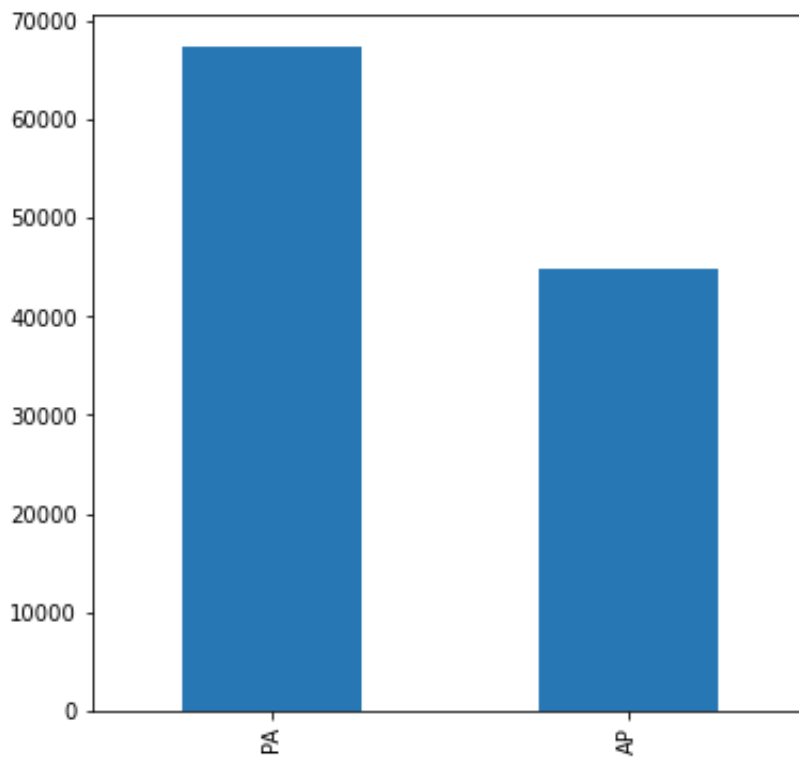
The gender distribution shows that there are more male than female figures (ratio 3:2 approximately).



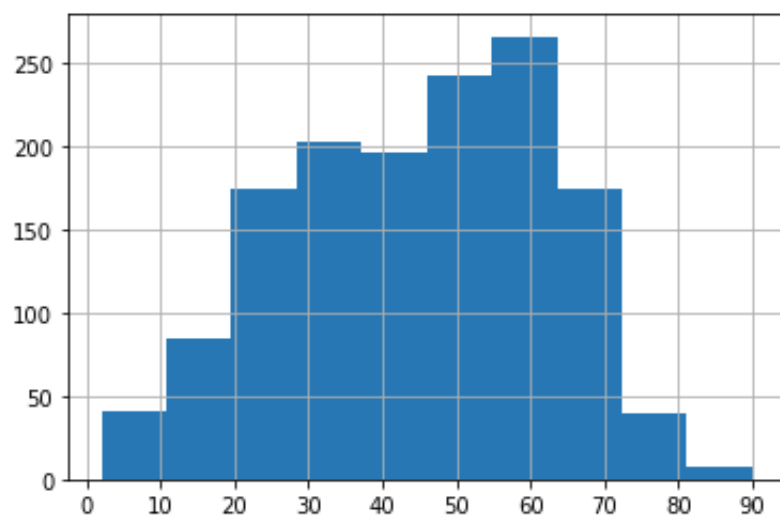
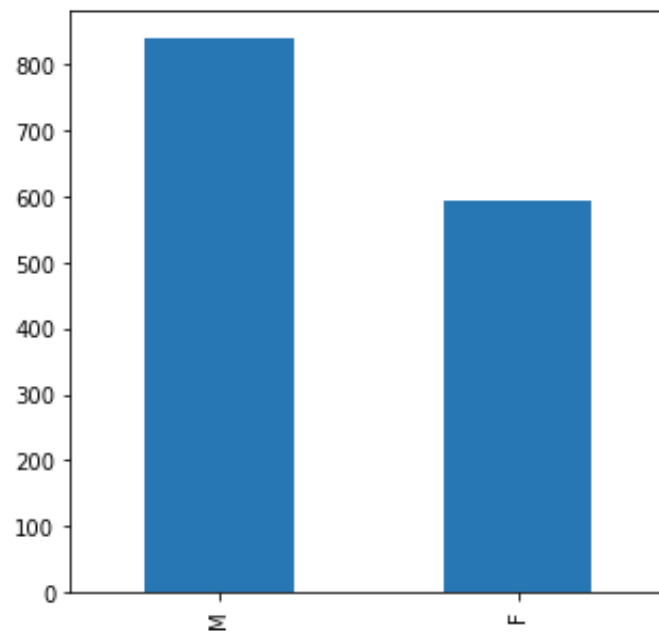
The whole dataset contains images from patients aging less than 90 years, but the bulk of patients covers 20–70 years..

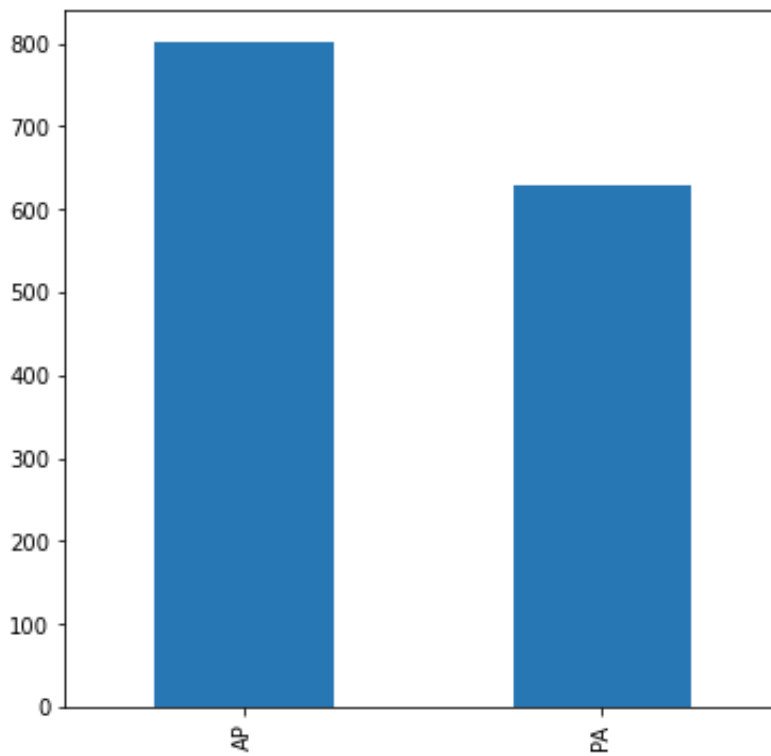


There are two possible patient positions: PA or AP. As we can see, the PA view dominates over the AP.

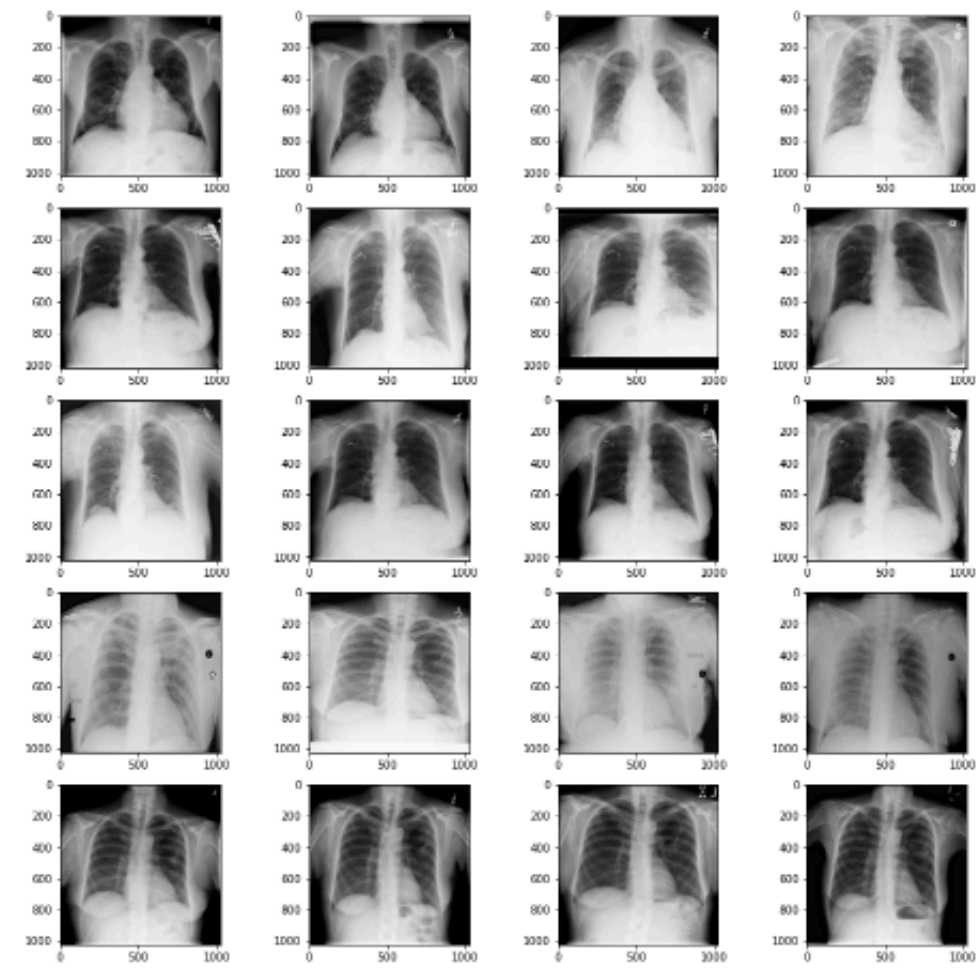


For patients containing Pneumonia, this graphs are as follows:





That is: the ratio of male:female images is approximately equal, patients with pneumonia are mostly aging 45–65 years and for pneumonia cases the dominant view is AP. An example of 20 images is shown below.



**Ground Truth Acquisition Methodology:** The NIH dataset radiologist reports were mined using Natural Language Processing (NLP) to create the disease labels from the associated radiology reports. The silver standard involves hiring several radiologists to each make their own diagnosis of an image. The final diagnosis is then determined by a voting system across all of the radiologists' labels for each image. Note, sometimes radiologists' experience levels are taken into account and votes are weighted by years of experience.

**Algorithm Performance Standard:** the threshold was chose to favour recall and then minimise the number of false negatives. The mean F1 score for radiologist is about 0.387 (95% CI 0.330, 0.442) [1], and this CNN scores a 0.36, which is pretty close to the standard of reference.

## References

[1] Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).