

# FDA Submission

---

**Your Name:** Cesar Gustavo Seminario Calle

**Name of your Device:** Identification of Pneumonia from X-Ray imaging

## Algorithm Description

### 1. General Information

**Intended Use Statement:**

For assisting the radiologist in the detection of Pneumonia using x-rays.

**Indications for Use:**

- Emergency workflow re-prioritization.
- This algorithm is intended for patients (women and men) between the ages of 20 to 60 years old whose X-ray has been taken in PA, AP position.
- The patients can have zero or more disease, specifically Atelectasis, Edema and Infiltration.

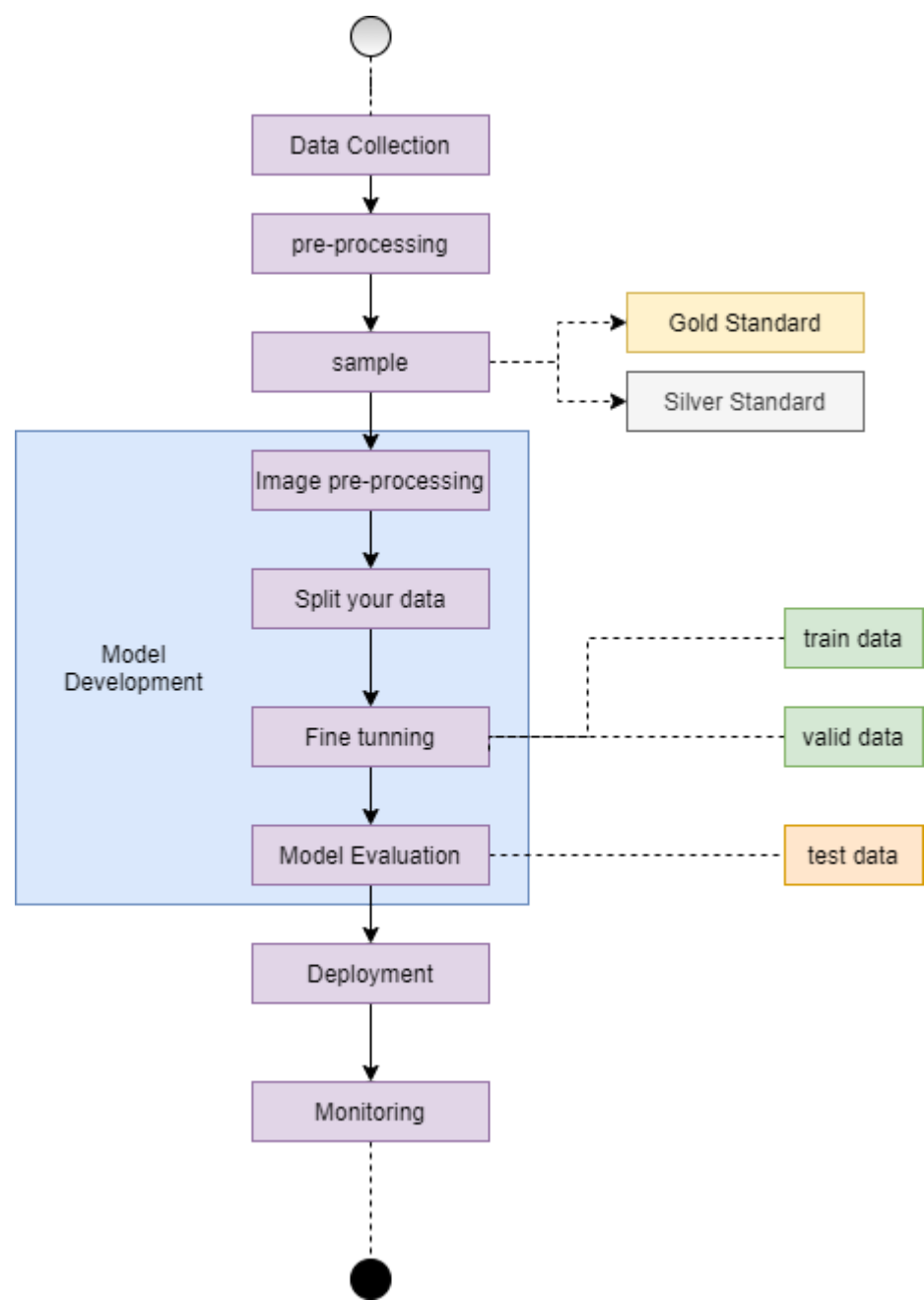
## Device Limitations:

- The algorithm will take more time at inference if it is not run in a GPU.
- Patients over 60 years old would cause the algorithm output not valid results.

**Clinical Impact of Performance:**

- A **false positive** will overcharge the priority list, making the radiologist focus on cases which don't need immediately attention.
- A **false negative** will have a high impact in the prioritization of the queue, putting patient in high risk in the final positions of the priority list.

### 2. Algorithm Design and Function



DICOM Checking Steps:

- The first step is to check if the patient position is PA or AP.
- The second step is to check the modality value is DX (Digital Radiography)
- The third step is to check that the body part examined is the chest part.

Preprocessing Steps:

1. The first preprocess step start dividing the image into 255 to scale the image
2. The next step involves extract the mean of the image from each pixel and divide into the standard deviation

CNN Architecture:

Model: "sequential_1"		
Layer (type)	Output Shape	Param #

```

=====
model_1 (Model)                (None, 7, 7, 512)                14714688
-----
flatten_1 (Flatten)            (None, 25088)                    0
-----
dense_1 (Dense)                (None, 512)                     12845568
-----
dropout_1 (Dropout)            (None, 512)                      0
-----
dense_2 (Dense)                (None, 256)                     131328
-----
dropout_2 (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

The following methods were used during training:

Method	Setting
rescale	1/255.0
horizontal_flip	No
vertical_flip	No
height_shift_range	0.05
width_shift_range	0.05
rotation_range	3
shear_range	0.9
zoom_range	0.1

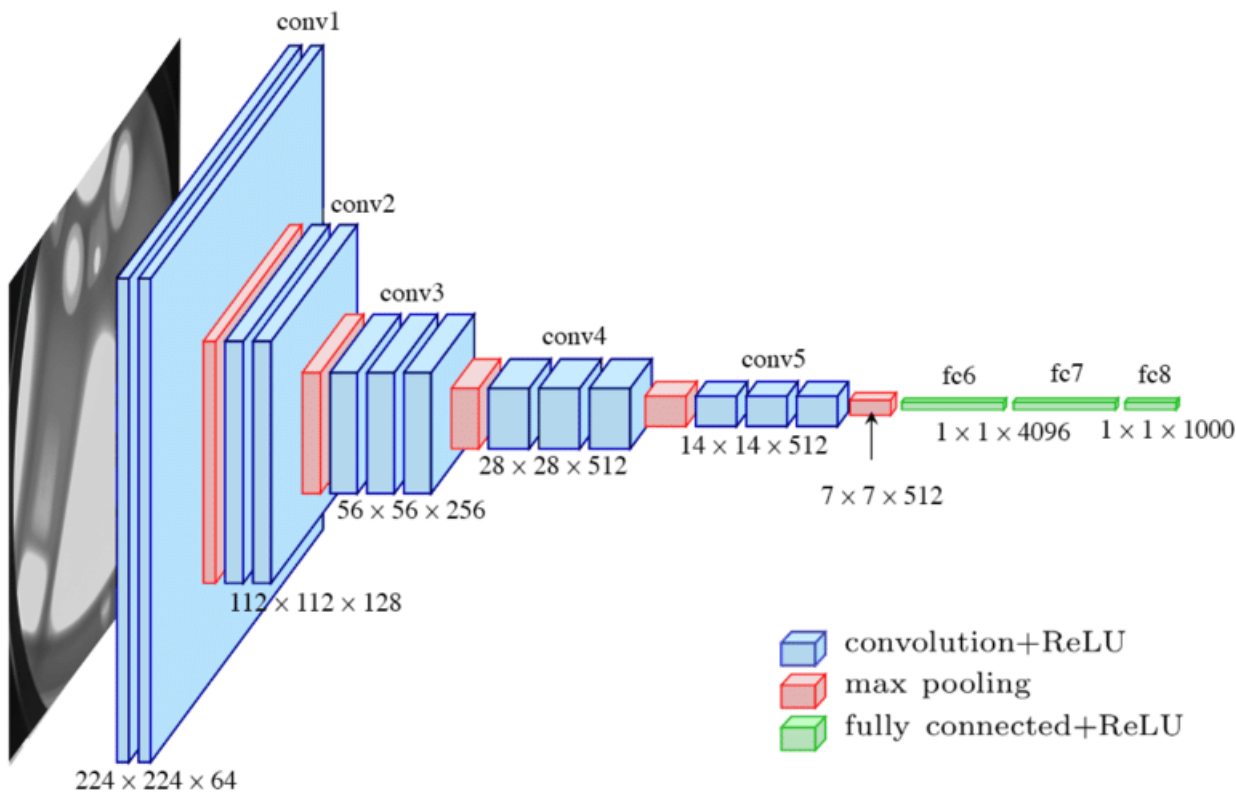
the rescale method was use to reduce the intensity of the image, there was no necessary to apply horizontal flip or vertical flip to the image due to images in a vertical position are not possible in a clinical context.

- Batch size  
16
- Optimizer learning rate  
0.001

- Layers of pre-existing architecture that were frozen

The first 17 layers of the pretrained model were frozen.

- Layers of pre-existing architecture that were fine-tuned



Model: "model\_1"

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
=====		

- Total params: 14,714,688
- Trainable params: 14,714,688
- Non-trainable params: 0

- Layers added to pre-existing architecture

```

model.add(Dense(512, activation='relu'))
model.add(Dropout(0.3))

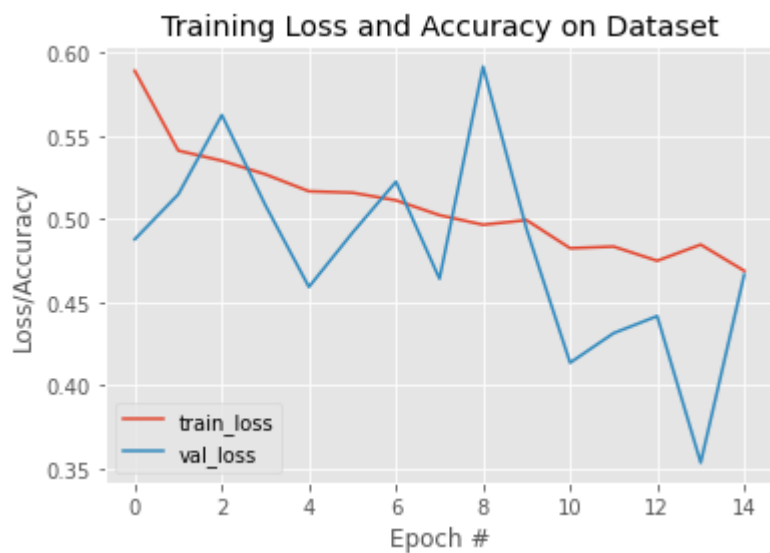
model.add(Dense(256, activation = 'relu'))
model.add(Dropout(0.2))

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

```

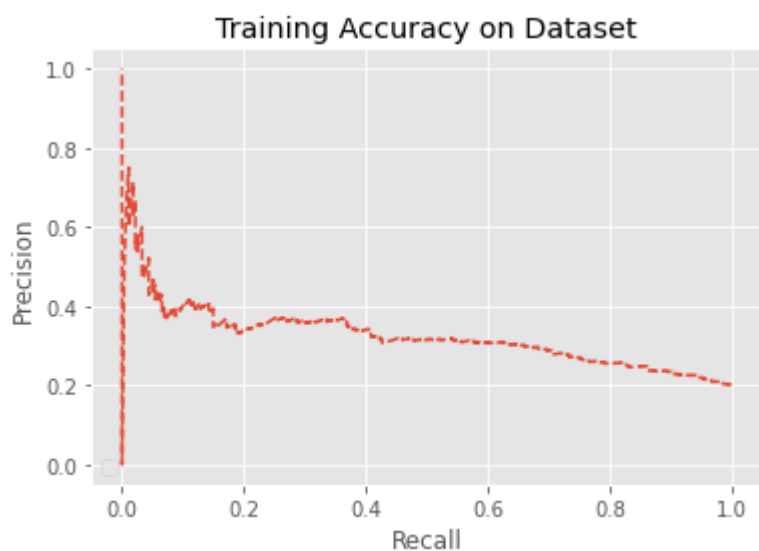
- Train/Validation Loss

The training loss decrease smoothly in each epoch, although the model was tested with different values of dropout and some layers, the validation loss was unstable.



- Precision Recall Curve

We can observe that middle point between precision and recall lays between 0.2 to 0.4 and 0.5 0.7 for recall and precision respectively.



#### Final Threshold and Explanation:

precision	recall	threshold	f1_score
0.307839388	0.587591241	<b>0.4285991</b>	<b>0.40201005</b>
0.2749658	0.733576642	0.3571264	0.400398406
0.32132964	0.423357664	0.49709988	0.362776025
0.202522255	0.996350365	0.03006885	0.336829118
0.407894737	0.113138686	0.78058845	0.17765043
0.418604651	0.065693431	0.85368896	0.113924051
0.425	0.062043796	0.85699904	0.108626198

precision	recall	threshold	f1_score
0.5	0.04379562	0.8939034	0.080808081
0.448275862	0.047445255	0.88102597	0.079470199
0.5	0.040145985	0.89512306	0.06779661
0.5625	0.032846715	0.9055565	0.062283737
0.5	0.032846715	0.90014625	0.06185567
0.5	0.03649635	0.89711297	0.061433447
0.571428571	0.02919708	0.9105697	0.048780488
0.625	0.018248175	0.95297354	0.035587189
0.666666667	0.02189781	0.9472421	0.035460993
0.714285714	0.018248175	0.9551629	0.028571429
0.666666667	0.01459854	0.964299	0.021505376
0.75	0.010948905	0.97217417	0.014440433
0.666666667	0.00729927	0.9725735	0.007246377
0.5	0.003649635	0.9916397	0
0	0	0.9945082	0

The threshold selected is 0.4285991, it give us a precision of  $\sim 0.3078$ , a recall of  $\sim 0.5875$  and a f1 score of  $\sim 0.40201$ .

#### 4. Databases

##### Description of Training Dataset:

The training dataset contains 1145 images randomly selected which are equally distributed accross each label (50% yes, 50 % no) for the presence of pneumonia.

Additionally, these images contains the presence of others diseases, those are the proportions with respect to the training dataset:

Disease	%
Infiltration	29.213974
Atelectasis	13.886463
Effusion	14.803493
Edema	12.358079
Consolidation	6.419214
Nodule	6.200873

Disease	%
Mass	5.152838
Pleural_Thickening	3.493450
Pneumothorax	3.624454
Cardiomegaly	2.576419
Emphysema	1.659389
Fibrosis	1.091703
Hernia	0.131004

**Description of Validation Dataset:** The validation dataset contains 248 images randomly selected which are distributed accross each label (20% yes, 80 % no) for the presence of pneumonia. Aditionally, these images contains the presence of others diseases, those are the proportions with respect to the validation data:

Disease	%
Infiltration	22.797203
Effusion	13.566434
Atelectasis	12.797203
Consolidation	4.825175
Edema	5.664336
Mass	5.384615
Nodule	5.454545
Pleural_Thickening	3.426573
Pneumothorax	4.195804
Cardiomegaly	1.748252
Emphysema	1.678322
Fibrosis	1.118881
Hernia	0.069930

## 5. Ground Truth

The dataset was labeled using Natural Language Processing from the associated radiological reports. The advantage of this method is that we can label many images in a short period of time. A representative sample of this dataset labeled can be contrasted against an specialist if neccesary. The model output expect to be 90% accurate, about 10% of the total labels are erroneous.

## 6. FDA Validation Plan



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

Persons distributed between the ages of 20 - 60. The images should be in the format of a Digital xray, which have taken in a PA or AP view position.

**Ground Truth Acquisition Methodology:**

The ground truth for the dataset can be achieved using a voting system of a team of radiologist assessment with different years of experience.

**Algorithm Performance Standard:**

		F1 score
Study 1	Radiologist 1	0.383
	Radiologist 2	0.352
	Radiologist 3	0.365
	Radiologist 4	0.442
	Radiologist Avg	0.387

The base performance metric is the f1-score which focus in the importance of balanced the proportion of false negatives and false positives that our models outputs. From the study *Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning* we define a f1-score of 0.387 as a baseline for our algorithm.