FDA Submission

Your Name: Aboubacar DIALLO

Name of your Device : Chest X-Rays Pneumonia Detector

Algorithm Description

1. General Information

Intended Use Statement : Assist the radiologist in the detection of Pneumonia with in Chest-X-Rays

Indication for Use : X-Ray image must be

- In DICOM format
- be taken in the AP or PA position
- For Male and Female
- Range of Age from 1 to 90 years old

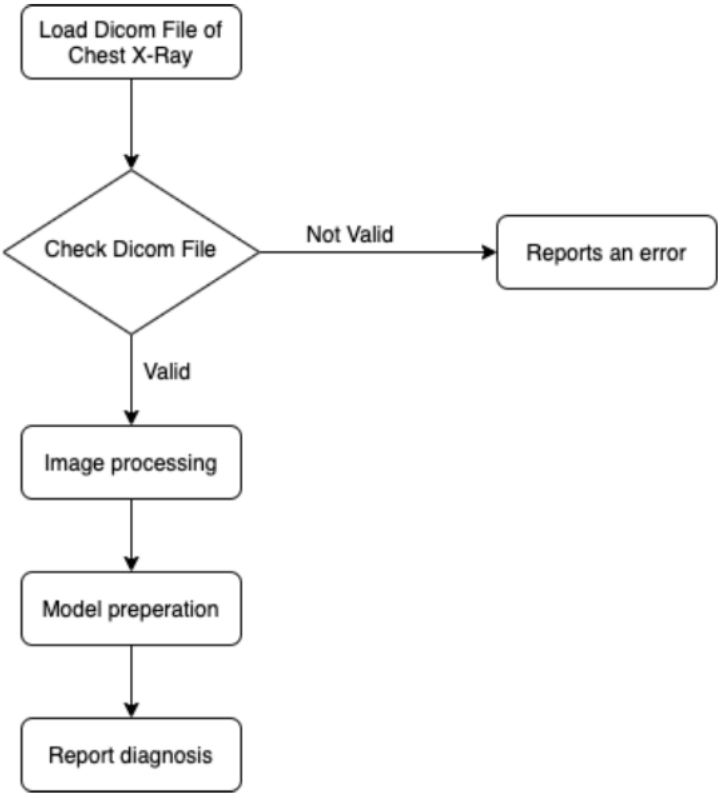
Devise Limitations :

Must be used with Computer with GPU for better accuracy

Clinical Impact of Performance :

- The model has a lower precision and higher recall
- Model most accurate when the test result is negative

2. Algorithm Design and Function



DICOM Checking Steps :

- Modality is DX
- CHEST is the mody part tobe examined
- Position PA or AP

Preprocessing Steps :

- Normalisation of Image
- Reshaping of Image
- Image is repeated across 3 channels

CNN Architecture :

- Model based on the VGG16 model
- The model uses the 16 years layers of the VGG16model

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
model_1 (Model)              (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              25691136
_________________________________________________________________
dropout_2 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 512)               524800
_________________________________________________________________
dropout_3 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_3 (Dense)              (None, 256)               131328
_________________________________________________________________
dropout_4 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 257
=================================================================
Total params: 41,062,209
Trainable params: 28,707,329
Non-trainable params: 12,354,880
```

3. Algorithm Training

Parameters

Image augmentation

```
idg = ImageDataGenerator(rescale = 1 / 255.0,
                         horizontal_flip = True,
                         vertical_flip = False,
                         height_shift_range = 0.1,
                         width_shift_range = 0.1,
                         rotation_range = 25,
                         shear_range = 0.1,
                         zoom_range = 0.15)
```

BATCH_SIZE=64
Learning = 1e-4

Final Threshold and Explanation

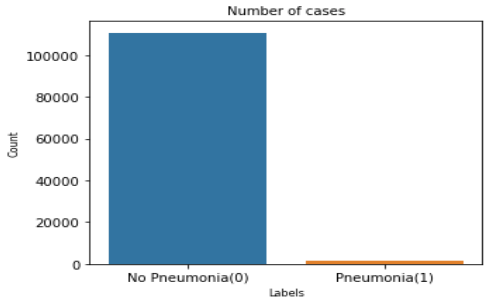
```
Precision is: 0.42857142857142855
Recall is: 0.9375
Threshold is: 0.35657388
F1 Score is: 0.588235294117647
```

4. DataBases

NIH Chest X-ray Dataset it contains 112,000 chest-ray images with disease labels acquired from from 30,000 patients
The dataset is composed :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 112120 entries, 0 to 112119
Data columns (total 12 columns):
 #   Column                       Non-Null Count   Dtype  
---  ------                       --------------   -----  
 0   Image Index                  112120 non-null  object 
 1   Finding Labels               112120 non-null  object 
 2   Follow-up #                  112120 non-null  int64  
 3   Patient ID                   112120 non-null  int64  
 4   Patient Age                  112120 non-null  int64  
 5   Patient Gender               112120 non-null  object 
 6   View Position                112120 non-null  object 
 7   OriginalImage[Width          112120 non-null  int64  
 8   Height]                      112120 non-null  int64  
 9   OriginalImagePixelSpacing[x  112120 non-null  float64
 10  y]                           112120 non-null  float64
 11  Unnamed: 11                  0 non-null       float64
dtypes: float64(3), int64(5), object(4)
memory usage: 10.3+ MB
```

```
0    110689
1      1431
Name: Pneumonia, dtype: int64
```



Description of Training Dataset :

Pneumonia and non-pneumonia patients are equally represented in the training data around 2290 images

Description of Validation Dataset
Pneumonia with patient is around 20% of the training, whereas non-pneumonia patients is around 80%

5. Ground Truth

The labels should be more than 90 correct and sufficient for weekly supervised learning

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset :

- Applicable to men and women from 1 to 90 years old
- DICOM format of chest X-rays images
- Must be Taken in the PA or AP position

Ground Truth Acquisition Methodology :

- Silver standard : Validation by more than 3 professional radiologists

Algorithm Performance Standard :

F1 score should be more than the average of radiologist (0.378)