FDA SUBMISSION

Masinde

2022-05-19

FDA Submission

Masinde Mtesigwa Masinde

Name of your Device: MassCancello

Algorithm Description

1. General Information

Intended Use Statement: This algorithm is intended for use on assisting radiologist in detecting pneumonia for the patients who have been administered a screening chest x-ray **Indications for Use:** MassCancello is an image processing software that provides qualitative and quantitative analysis of the chest from x-ray images to support clinicians in the evaluation and assessment of pneumonia disease.

Device Limitations: The results above indicate that the presence of infiltrations in a chest x-ray is a limitation of this algorithm, and that the algorithm performs very poorly on the accurate detection of pneumonia in the presence of infiltration.

Clinical Impact of Performance: The presence of nodules and pneumothorax have a some effects on the algorithm's sensitivity and which might reduce the ability to detect pneumonia, and the presence of effusion has a slight impact on specificity and may increase the number of false positive pneumonia classifications.

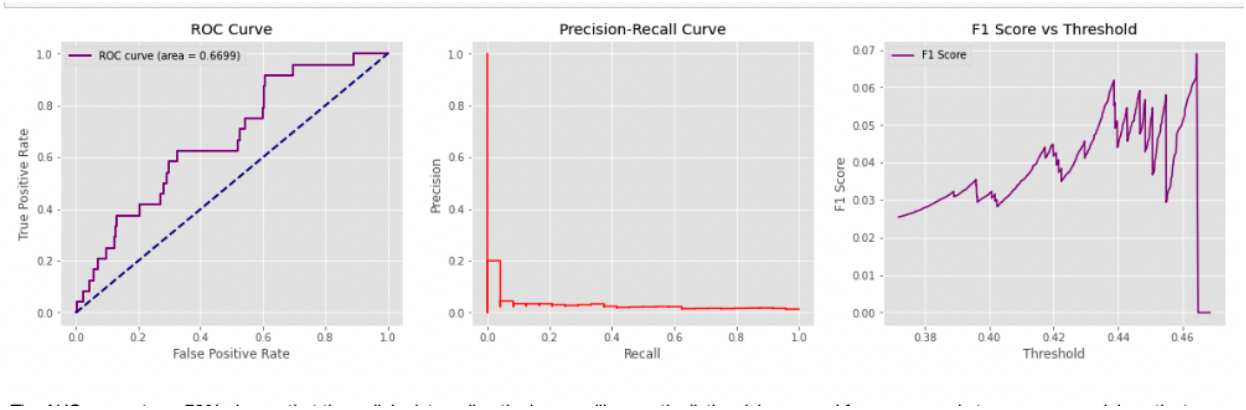

Figure 1: PR ROC F1

2. Algorithm Design and Function

DICOM Checking Steps: The first step is pre-extract all data from DICOM headers into a dataframe - Body Part Examined - A patient age - A patient ID - Patient Sex (M/F) - DICOM Study number

Preprocessing Steps:

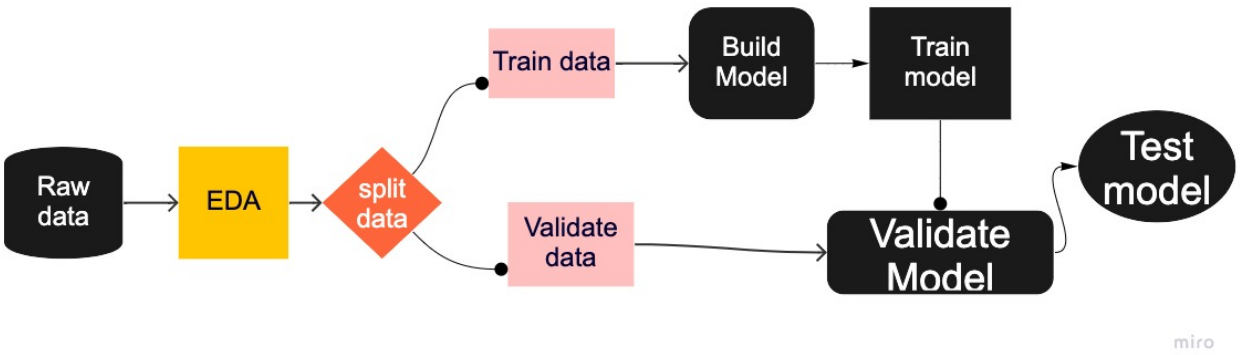

Figure 2: The process flowchart

CNN Architecture: The CNN algorithm which is used is VGG16

3. Algorithm Training

Parameters:

- Types of augmentation used during training
  - directory=None
  - x_col = 'path'
  - y_col = 'pneumonia_class'
  - class_mode = 'binary'
  - target_size =IMG_SIZE
  - batch_size = 16
- Batch size = 16 and 2000
- Optimizer learning rate = Adam(lr=1e-4)
- Layers of pre-existing architecture that were frozen [0:-2]
- Layers of pre-existing architecture that were fine-tuned
- Layers added to pre-existing architecture

Performance

P-R curve

The figure 4 shows PR curve

Final Threshold and Explanation: With the threshold of 0.65 shows that the image has pneumonia with reference to figure 4 ### 4. Databases

Description of Training Dataset: Data set have

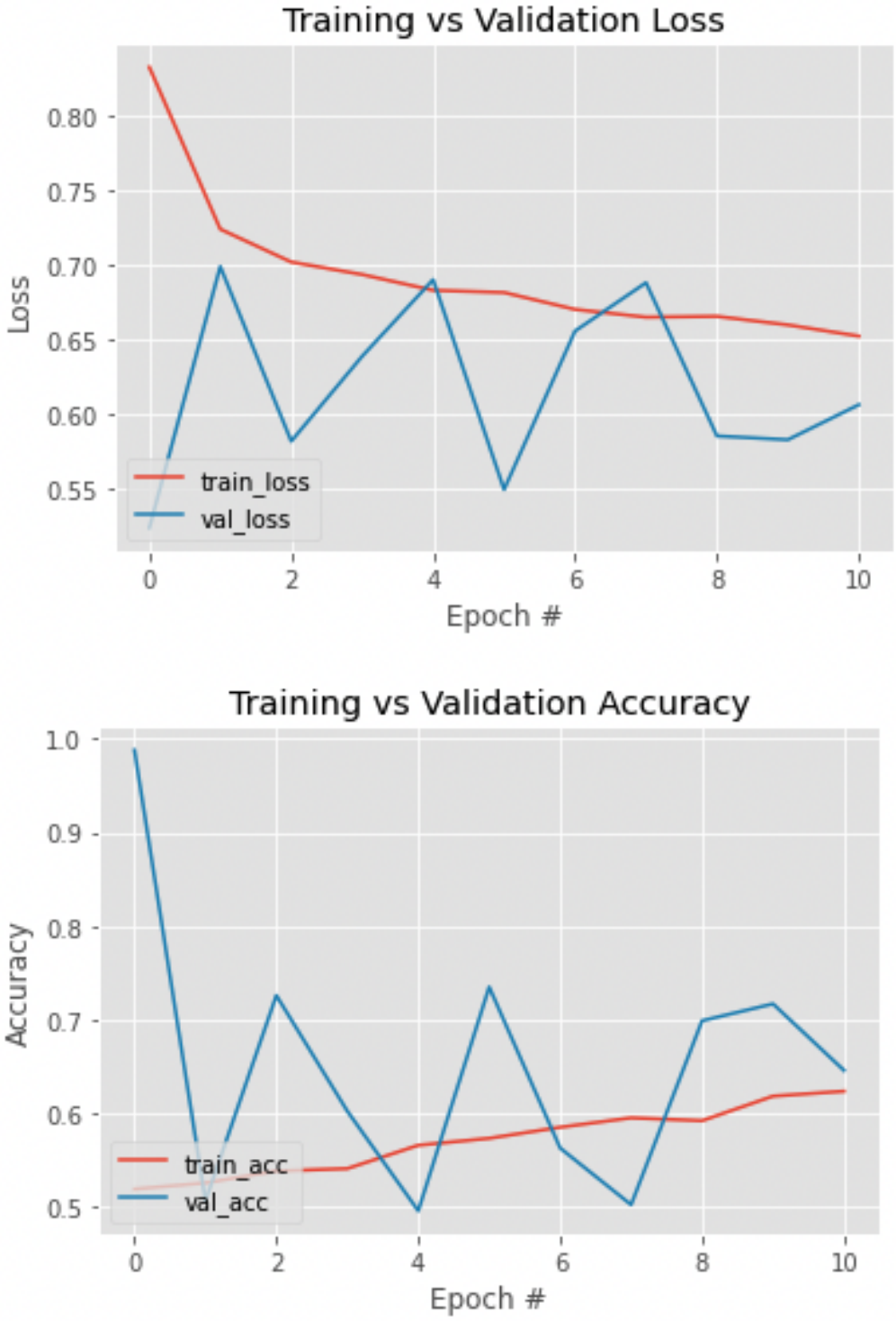

Figure 3: Algorithm performance

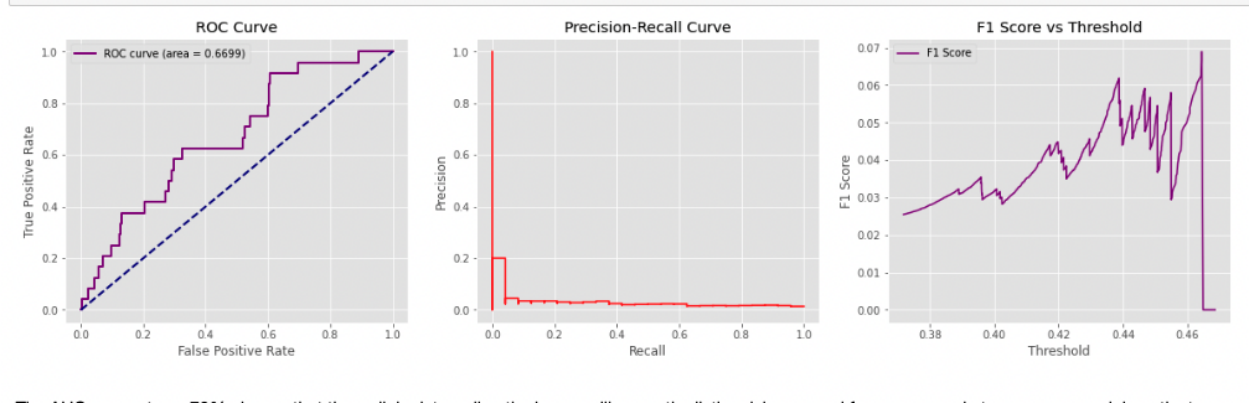

Figure 4: PR curve

Pneumonia cases: 1,430 Non-Pneumonia cases: 110,674

Our training set to be balanced between Pneumonia and Non-Pneumonia cases Our test set to reflect the real world proportions To split our data between training and test sets in a 80% to 20% proportion

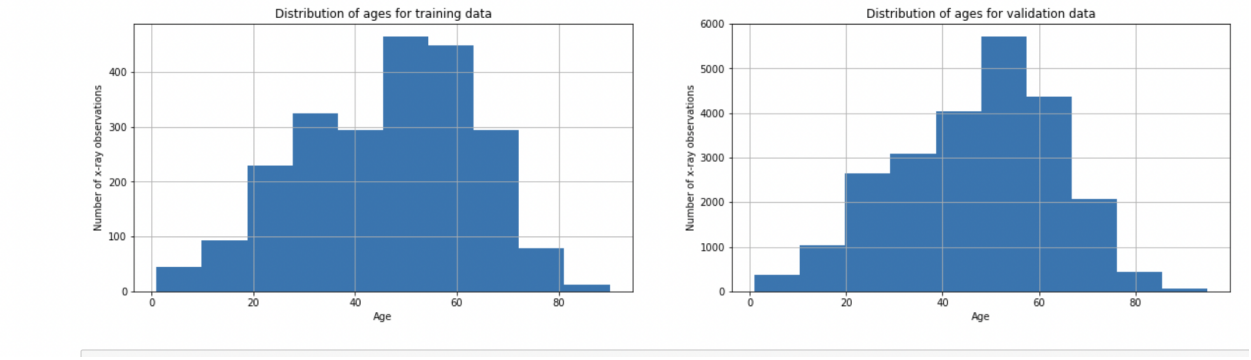

Figure 5: Distributions of training vs validation data

Description of Validation Dataset:

5. Ground Truth

The radiologists labeling.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: The histograms shows that the age distribution that the patients are from age 5 to 95. The gender histogram also shows that most patients are male.

Description of Validation Dataset:

Ground Truth Acquisition Methodology: The ground truth method The silver standard approach of using several radiologists would be more optimal for this algorithm.

Algorithm Performance Standard: The algorithm does not perfom well this is because the accuracy is only 60%.

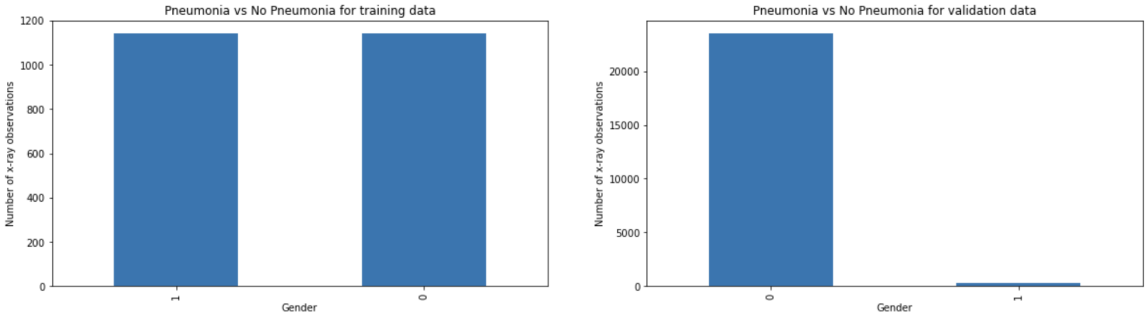

Figure 6: Pneumonia vs No Pneumonia cases between training vs validation data

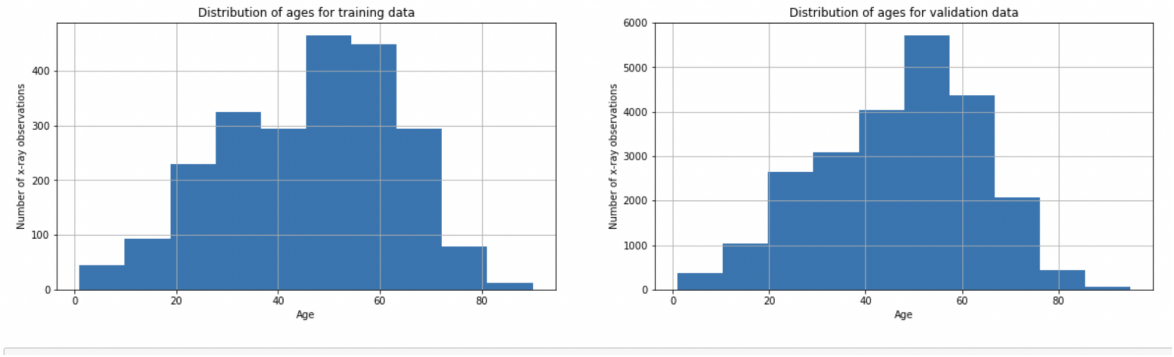

Figure 7: Distributions of training vs validation data