FDA Submission

Your Name: Sylwia Nowakowska

Name of your Device: Deep Learning algorithm for Pneumonia detection from chest X-rays

Algorithm Description

1. General Information

Intended Use Statement:
The deep learning algorithm detecting pneumonia from Chest X-ray scans taken in PA or AP viewing position, is intended to assist radiologists and should be used together with patient's medical history and current symptoms description. It is intended to be used for male and female patients aged 1 – 90.

Indications for Use:
Emergency workflow re-prioritization in Emergency Department on the basis of Chest X-ray scans taken in PA or AP viewing position, for male and female patients aged 1 – 90.

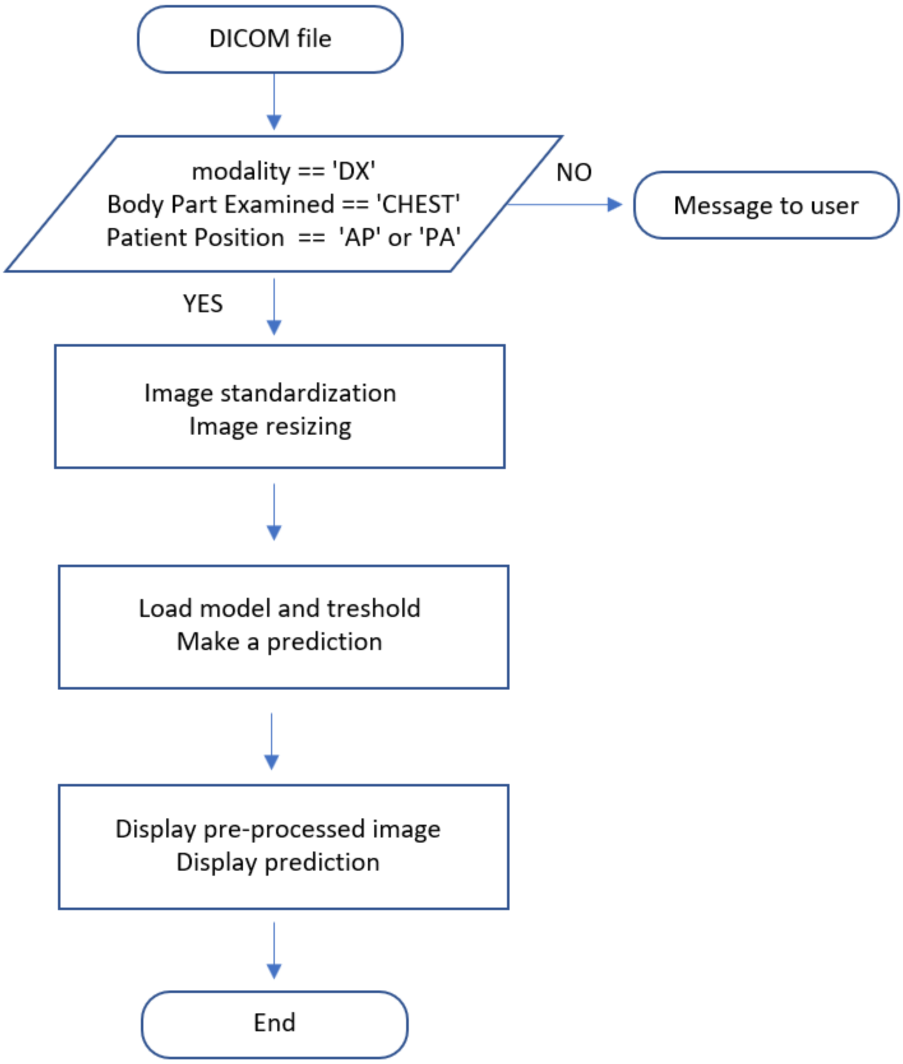
Device Limitations:
The algorithm must be run on a computer meeting minimum requirements of GPU and RAM.

Clinical Impact of Performance:
Radiologists using the algorithm are required to review the Chest X-ray scans and to validate the algorithm prediction taking into account patient's medical history and current symptoms.

- If the device predicts False Positive the patient will have follow-up diagnostics tests, e.g. blood test, pulse oximetry. In case of serious symptoms blood test, sputum test, PCR test, CT or bronchoscopy.
- False negative result lead to patient's discharge and carries the risk of worsening of the patient state.

2. Algorithm Design and Function

Algorithm flowchart:



Flowchart 1 Algorithm flowchart

DICOM Checking Steps:
Each DICOM file's header is checked, whether:

- the modality is 'DX'
- the Body Part Examined is 'CHEST'
- the Patient Position is 'AP' or 'PA'

Preprocessing Steps:
If the DICOM file passes the checking step, image pre-processing takes place:

Step 1)

Image standardization: A representative mean and standard deviation was calculated from training batch consisting of 100 scans. From each pixel of the image to be analysed this mean value is subtracted and the resulting value is divided by the standard deviation. This assures proper functioning of the algorithm, which was trained on data pre-processed in the same way.

Step 2)

Image resizing: The image is resized to the size of 224 pixels by 224 pixels, which is the size of the input required by the algorithm.

CNN Architecture: A sequential model was built by using pre-trained model DenseNet121 with imagenet weights. Following layers were attached to its output:

- Dense Layer with 500 neurons and ReLu activation function
- Dropout layer with dropout value of 0.2
- Dense Layer with 100 neurons and ReLu activation function
- Dropout layer with dropout value of 0.2
- Dense Layer with 50 neurons and ReLu activation function
- Dense Layer with single neuron and Sigmoid activation function being an output layer

The last but one layer was set to transfer layer, i.e. from this layer the weights were adjusted during training. This assures that the pre-trained model preserves its ability to recognize general features, while specialises on Chest X-ray for pneumonia detection.

3. Algorithm Training

Parameters:

- Types of augmentation used during training:
  ImageDataGenerator from Keras library was used to perform data augumentation.
  Following augumentation was performed:
  - random shifting by 0.1 fraction of total height (height_shift_range=0.1)
  - random shifting by 0.1 fraction of total width (width_shift_range=0.1)
  - random rotations by +-2 deg (rotation_range=2)
  - random zooming by 0.01 fraction (zoom_range=0.01)

- Batch size
  The batch size of the training, validation and test data was set to 64.

- Optimizer learning rate:
  The starting rate of the Adam optimizer was set to 0.0001.

- Layers of pre-existing architecture that were frozen:
  All layers of the DenseNet121 model up to the last but one layer were frozen, i.e. their weights were not adjusted during training.
- Layers of pre-existing architecture that were fine-tuned:
  Following layers of the DenseNet121 model were fine tuned:
  - avg_pool
  - fc1000

- Layers added to pre-existing architecture
  Following layers were added to the DenseNet121 model:
  - Dense Layer with 500 neurons and ReLu activation function
  - Dropout layer with dropout value of 0.2
  - Dense Layer with 100 neurons and ReLu activation function
  - Dropout layer with dropout value of 0.2
  - Dense Layer with 50 neurons and ReLu activation function
  - Dense Layer with single neuron and Sigmoid activation function being an output layer

Training and validation:

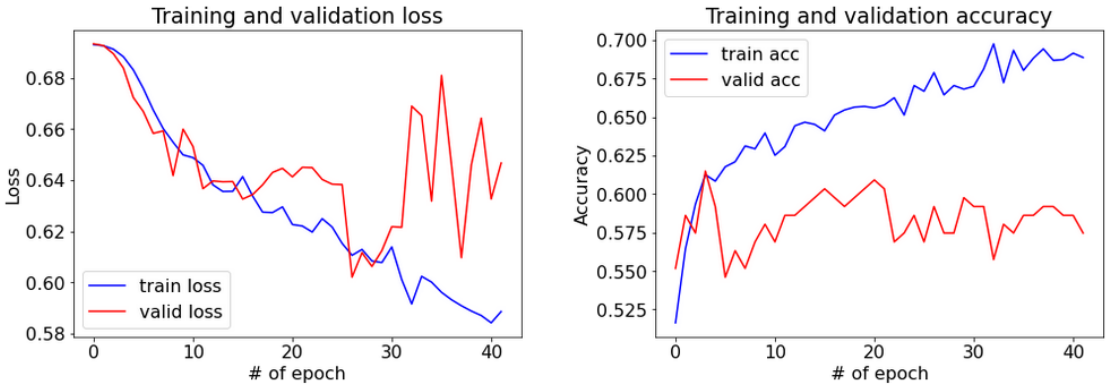


Figure 1 Algorithm performance during training and validation

During training, the training and validation loss and accuracy were monitored. In case the validation loss did not decrease throughout 15 epochs, the training was stopped. This was the case for the model described here, the model weights were saved after 27th epoch. The training loss (accuracy) was decreasing (increasing) throughout the algorithm learning. The validation loss was decreasing till circa 15 epoch, afterwards it was very noisy. The validation accuracy was increasing until 5th epoch, afterwards it was very noisy and lower than training accuracy.

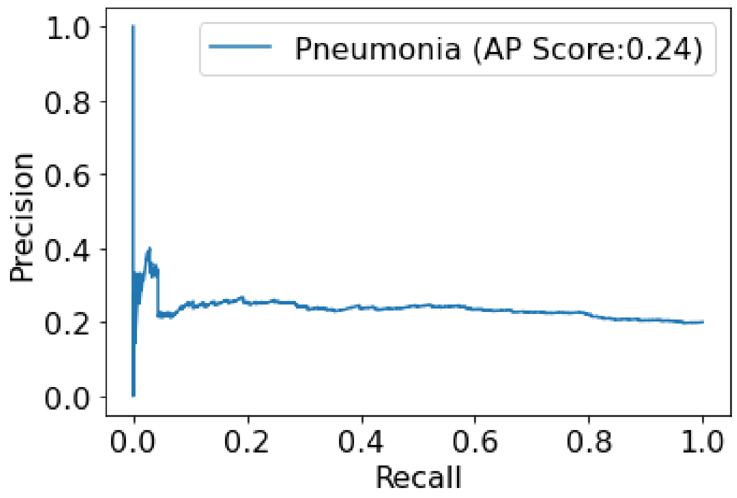


Figure 2 Algorithm precision-recall curve

The precision is high for low recall values. For higher recall values the precision stabilizes.

Final Threshold and Explanation:

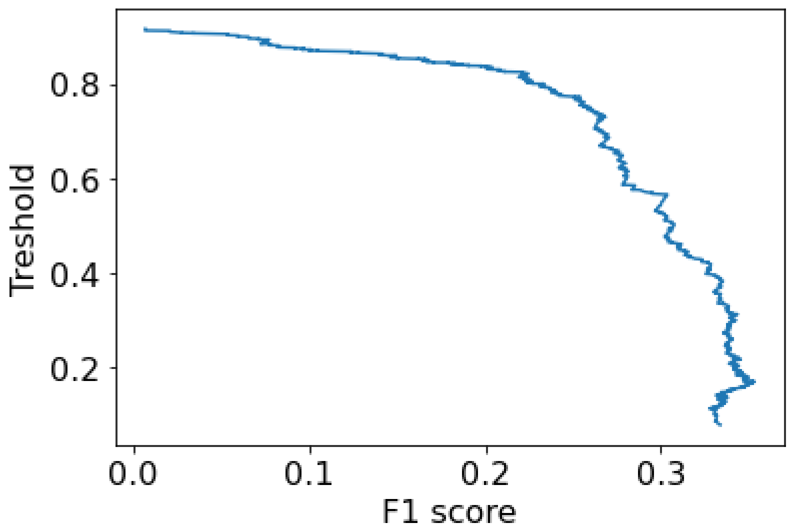


Figure 3 Algorithm treshold vs F1 score curve

A threshold value of 0.171, maximizing the F1 score was chosen on the basis of the plot above. This threshold value resulted in following metrics:

- F1 score = 0.353
- Precision = 0.228
- Recall 0.785.

4. Databases

NIH Chest X-ray Dataset comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. Detailed info: https://www.kaggle.com/nih-chest-xrays/data

The training, validation and test datasets were created in such a way that there is no patient overlap between them. The training and validation sets were balanced, whereas the test set was imbalanced, such that the prevalence of pneumonia cases resembles the prevalence in an Emergency Department.

Description of Training Dataset:

- Number of scans: 2140
- Number of pneumonia cases: 1070 (constituting 50% of the dataset)

Description of Validation Dataset:

- Number of scans: 174
- Number of pneumonia cases: 87 (constituting 50% of the dataset)

Description of Training Dataset:

- Number of scans: 1370
- Number of pneumonia cases: 274 (constituting 20% of the dataset)

Patient demographic data

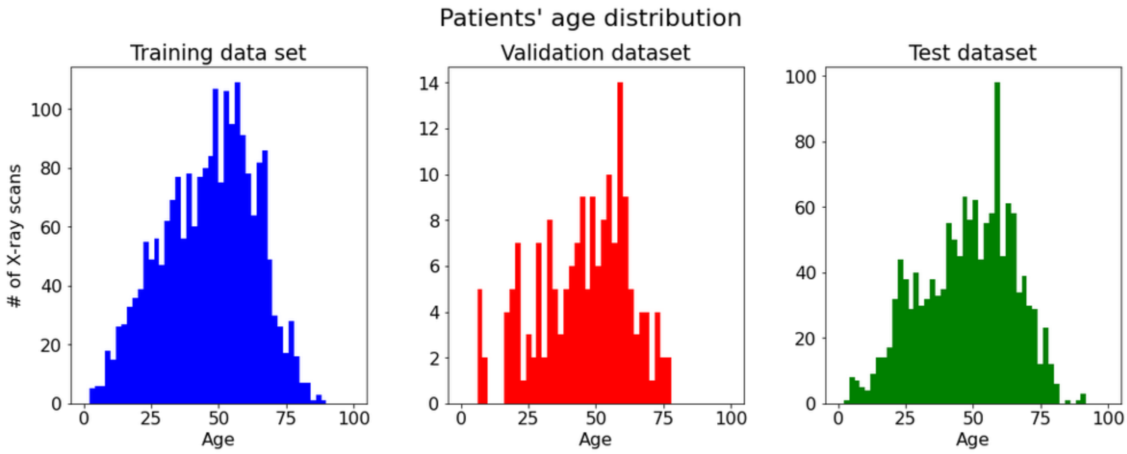


Figure 4 Patients' age distribution in the datasets

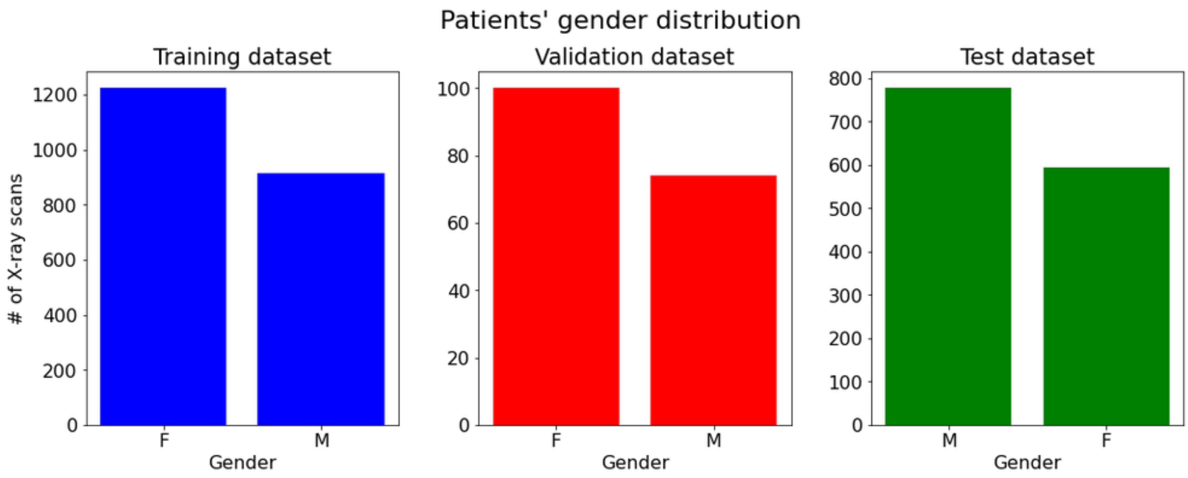


Figure 5 Patients' age distribution in the datasets

Patient view positions

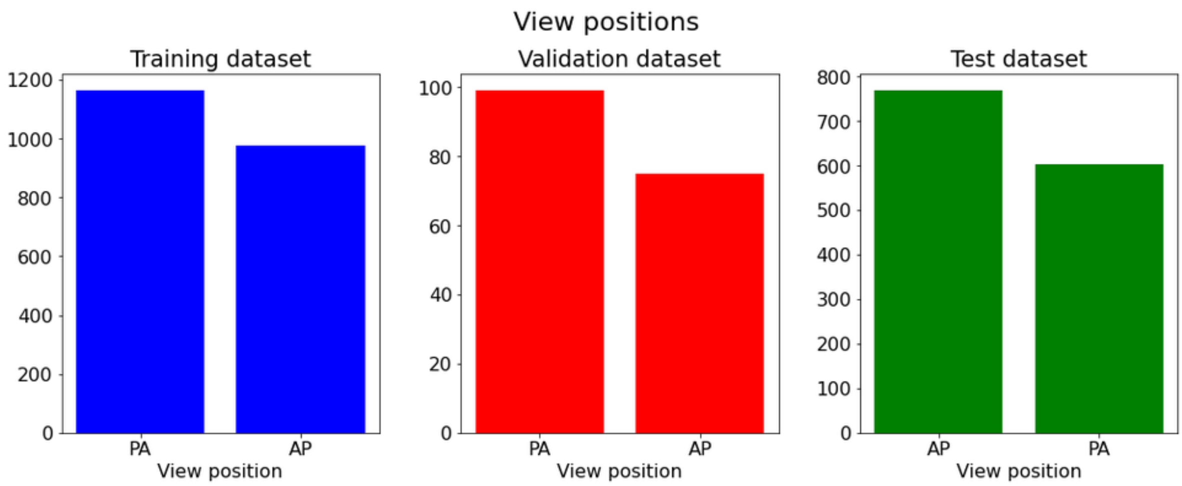


Figure 6 View position distribution in the datasets

Pneumonia and co-occurrent diseases and findings

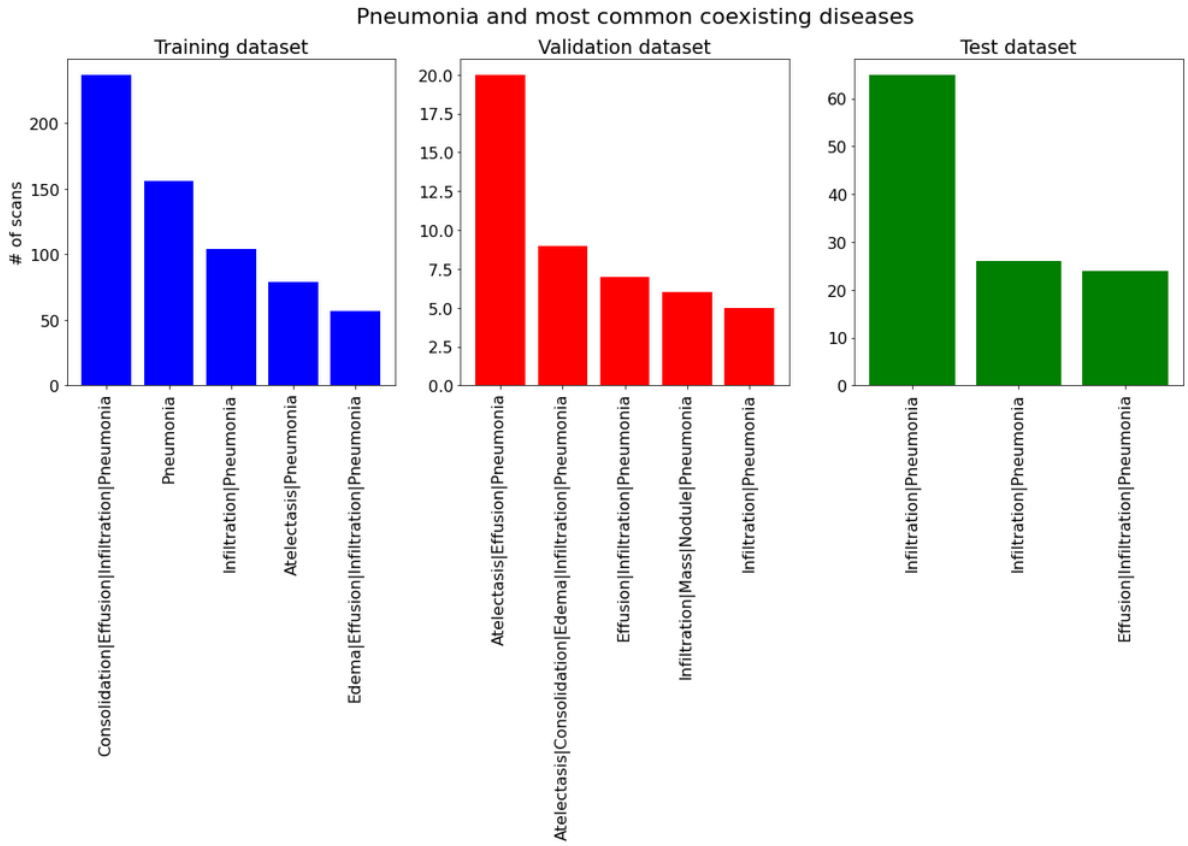


Figure 7 Pneumonia and most common coexisting diseases and findings in the datasets

5. Ground Truth

The dataset labels include 14 pathologies and were extracted using Natural Language Processing (NLP) from radiological reports. The benefit of such ground truth extraction is its automatization and speed, the drawback is that NLP not always extracts the right label, especially when it is embedded in complex sentences. Moreover, radiologists are not always diagnosing a patient correctly or are advising follow-up studies.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

- Age: 1-90 years old
- Sex: Male and Female
- Modality: Digital chest X-ray
- Body part imaged: Chest
- Prevalence of pneumonia: 20% in emergency setting

- No inclusion / exclusion of comorbidities

Ground Truth Acquisition Methodology:

- Consensus voting of 5 radiologist
- Conformed with follow-up PCR test

Algorithm Performance Standard:

The algorithm has F1 score amounting to 0.353, which is comparable to Radiologist 2 from Ref [1], who had F1 score 0.356.

[1] Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, et al. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." ArXiv:1711.05225 [Cs, Stat], December 25, 2017