FDA Submission

Chemseddine MENAKBI

NN-4-Pneumonia

Algorithm Description

1. General Information

Intended Use Statement*:* Pneumonia is characterized by white spots on xrays radiography. NN-4-Pneumonia is notification-only intended to produce a label "Pneumonia" or "non-pneumonia". The device is intended to assist hospital networks and trained radiologists for worklist prioritization or triage by flagging and communicating of suspected positive findings of pneumonia.

Indications for Use: The device uses an artificial intelligence algorithm to analyze normal postero-anterior (PA) and antero-posterior (AP) chest xrays in DICOM image data for features suggestive of pneumonia. NN-4-pneumonia can run on male and female patients images of any age.

Device Limitations: NN-4-Pneumonia is limited to analysis of imaging data as a guide to possible presence of pneumonia. The device is not intended to direct attention to specific portions or anomalies of an image. The device is not intended to be used for diseases other than pneumonia. The device does not replace review and diagnosis of the X-rays by radiologists and should not be used in lieu of full patient evaluation or relied upon to make or confirm diagnoses DICOM embedded image modality must be 'DX', body part examined must be 'CHEST', 'RIBCAGE', 'LUNG' or 'CHESTABDOMEN'. The device can be slow at suggesting the presence or absence of pneumonia on DICOM images data without the appropriate computer hardware and the use of a GPU. DICOM Image photometric Interpretation must be 'MONOCHROME1' or 'MONOCHROME2'.

Clinical Impact of Performance: The f-score of this device is 0.57 and the gmean is 0.73. Although this device is not intended to replace a radiologist, a false negative would imply that a patient with pneumonia has not been suggested as such and this should be avoided for the patient. A false positive would imply that a patient without pneumonia has been suggested as having it. This may lead to more analysis that would cost money and time for the patient but also a loss of radiologists and other physicians availability for other patients. False positive rate can be sacrificed a bit to decrease false negative rate if needed.
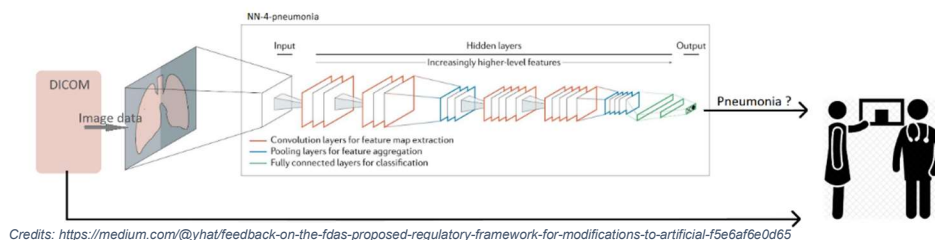
Substantial equivalence: NN-4-pneumonia has very similar intended use and indications, technological characteristics, and principles of operation as the AIMI-Triage CXR PTX. The minor differences in intended use and indications do not affect its safety and effectiveness when used as labeled. In addition, the minor technological differences between the device and its predicate raise no new issues of safety or effectiveness.

	NN-4-Pneumonia	AIMI-Triage CXR PTX
Intended Use / Indications for Use	NN-4-Pneumonia is notification-only intended to produce a label "Pneumonia"	The AIMI-Triage CXR PTX Application is a notification-only triage workflow tool for

	or "non-pneumonia". The device is intended to assist hospital networks and trained radiologists for worklist prioritization or triage by flagging and communicating of suspected positive findings of pneumonia. The device uses an artificial intelligence algorithm to analyze normal postero-anterior (PA) and antero-posterior chest x-ray (AP) embedded in DICOM image data for features suggestive of pneumonia. NN-4-pneumonia can run on male and female patients of any age. NN-4-Pneumonia is limited to analysis of imaging data as a guide to possible presence of pneumonia. The device is not intended to direct attention to specific portions or anomalies of an image. The device is not intended to be used for diseases other than pneumonia. The device does not replace review and diagnosis of the X-rays by radiologists and should not be used in lieu of full patient evaluation or relied upon to make or confirm diagnoses DICOM embedded image modality must be 'DX', body part examined must be 'CHEST', 'RIBCAGE', 'LUNG' or 'CHESTABDOMEN'. The device can be slow at suggesting the presence or absence of pneumonia on DICOM images data without the appropriate computer hardware and the use of a GPU. DICOM Image photometric Interpretation must be 'MONOCHROME1' or 'MONOCHROME2'.	use by hospital networks and clinics to identify and help prioritize chest X-rays acquired in the acute setting for review by hospital radiologists. The device operates in parallel to and independent of standard of care image interpretation workflow. Specifically, the device uses an artificial intelligence algorithm to analyze images for features suggestive of moderate to large sized pneumothorax; it makes case-level output available to a PACS/workstation for worklist prioritization or triage. Identification of suspected cases of moderate to large sized pneumothorax is not for diagnostic use beyond notification. The AIMI-Triage CXR PTX Application is limited to analysis of imaging data as a guide to possible urgency of adult chest X-ray image review, and should not be used in lieu of full patient evaluation or relied upon to make or confirm diagnoses. Notified radiologists are responsible for engaging in appropriate patient evaluation as per local hospital procedure before making care-related decisions or requests. The device does not replace review and diagnosis of the X-rays by radiologists. The device is not intended to be used with plain film X-rays.
User	Radiologist	

Anatomical Region	Lung	
Clinical Condition	Pneumonia	Pneumothorax
Modality	Chest X-ray	
Segmentation of region of interest	No; device does not mark, highlight, or direct users' attention to a specific location in the original image	
Alteration of original image	No	
Relation to standard of care workflow	Independent/parallel; no cases are removed from worklist queue	
Algorithm	Artificial intelligence algorithm with database of images	
Notification / Prioritization	Yes	
Alert to Finding	Passive notification – flagged for review	
Where Results are Received	PACS / Workstation	

2. Algorithm Design and Function

Credits: https://medium.com/@yhat/feedback-on-the-fdas-proposed-regulatory-framework-for-modifications-to-artificial-f5e6af6e0d65

DICOM Checking Steps: The device checks that the DICOM embedded image : modality must be 'DX', body part examined must be 'CHEST', 'RIBCAGE', 'LUNG' or 'CHESTABDOMEN'. NN-4-Pneumonia resizes images to 224x224 resolution, normalizes values (by dividing by 255) and convert grayscale DICOM images to 3 channels RGB.

CNN Architecture: NN-4-Pneumonia uses a sequential architecture of a VGG16 architecture. The dense layers are replaced by 4 dense layers of 1024, 512, 256, and 1 node respectively

# 3. Algorithm Training

Parameters:

- Types of augmentation used during training : random rotation (10°), random horizontal flipping, height shift (0.1), width shift (0.1), shearing (0.1), zoom (0.5, 1.0)
- Batch size : 128
- Layers of pre-existing architecture that were frozen : 17 first layers (input layer counts as the first layer)
- Layers of pre-existing architecture that were fine-tuned : last 2 convolutional layers

- Layers added to pre-existing architecture : 4 dense layers of 1024, 512, 256, and 1 node respectively

The training has been done with different learning rates and dropout probabilities for the 4 last dense layers:

- 20 epochs : dropout probability = 0.5, learning rate = 0.0001
- 60 epochs : dropout probability = 0.2, learning rate = 0.0001
- 10 epochs : dropout probability = 0.2, learning rate = 0.00001

The resulting performance metrics for the three training runs are presented in the following graphs.
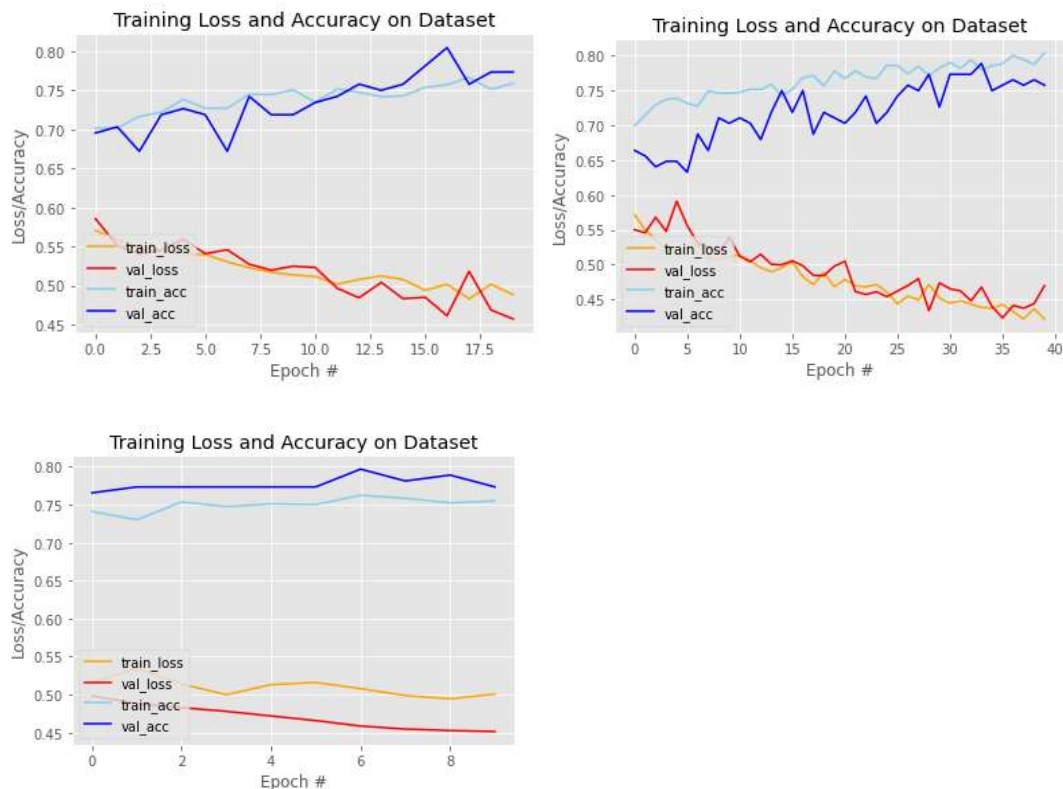
The final binary accuracy on the training set was : 0.76 and the final accuracy on an independent validation dataset of 512 images was: 0.80

The training resulted in the following p-r curve
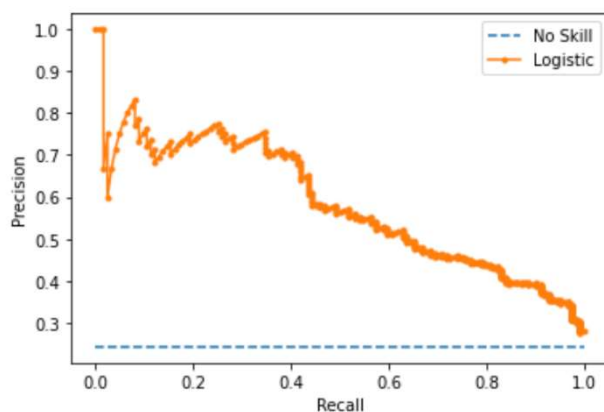
Final Threshold and Explanation: False negatives may be dangerous for the patient A higher precision would decrease false negatives rate.
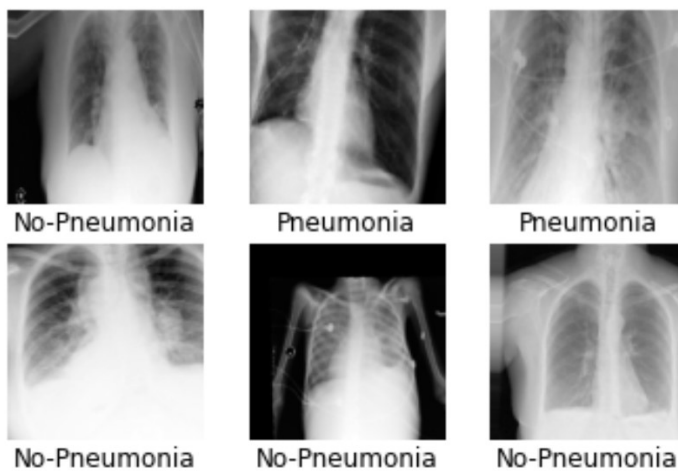
False positives may cost money and time for the patient. False positives also costs time for radiologists and other physicians that loose availability, which can be problematic for other eventual patients. A higher recall would imply less false positives.

The choice made for NN-4-pneumonia concerning the final threshold chosen is to give a similar importance to precision and recall and thus to choose a final threshold:0.44 that maximizes f1-score : 0.57 (f beta =1).

This is a first default choice that can be tweaked (choosing a different beta value for the f score leading to a different threshold) in collaboration with the radiologist according to his expertise and preferences and his consideration of the relative importance of false negatives and false positives.

4. Databases

Description of Training Dataset: The device was built and evaluated on an initial dataset presenting 112120 images from the chest xray NIH dataset used by Rajpurkar and al. [arXiv:1711.05225] downloadable on kaggle https://www.kaggle.com/nih-chest-xrays/data.

No-Pneumonia   Pneumonia   Pneumonia

No-Pneumonia   No-Pneumonia   No-Pneumonia

These 112120 images comesfrom 30805 unique patients. 51759 cases present a finding but only 1431 cases present pneumonia (1.27%). 13 other diseases (eventually pneumonia comorbidities) are present : 'Atelectasis', 'Cardiomegaly', 'Consolidation', 'Edema', 'Effusion', 'Emphysema', 'Fibrosis', 'Hernia', 'Infiltration', 'Mass', 'Nodule', 'Pleural_Thickening', 'Pneumothorax'.
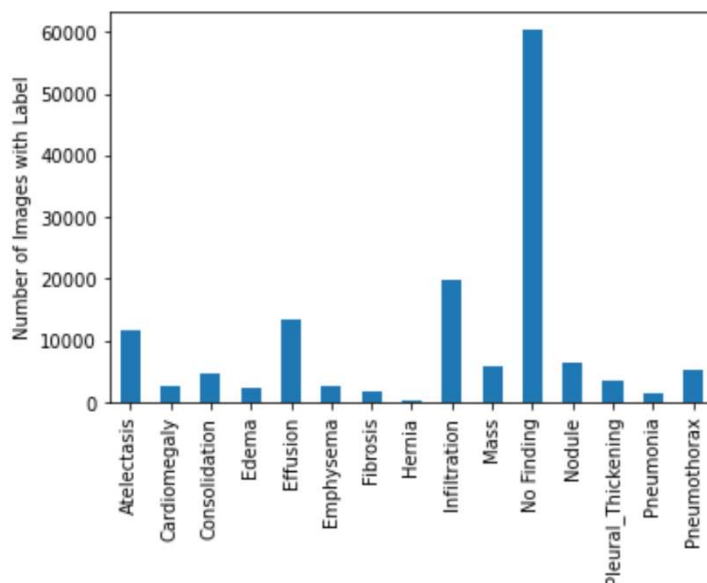
The average age of the whole dataset is 46.88 years old with a standard deviation of 16.60. 56.49% of the cases are male and 43.51% are female. The proportion of PA and AP view position is 60.03% and 39.97% respectively.

Cases presenting pneumonia have very similar statistics. The average age of pneumonia cases is 44.67 years (STD=17.6). Male, female proportions are 58.56% and 41.44% respectively. PA and AP view position proportions are 44.02% and 55.97%.

Total and pneumonia cases age distribution
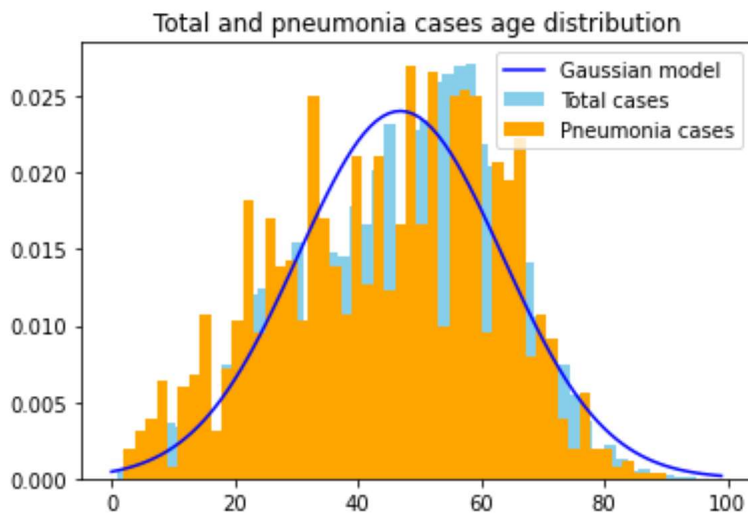
The distribution of diseases among the total initial dataset is :

Atelectasis   10.30%
Cardiomegaly   2.47%
Consolidation   4.16%
Edema   2.05%
Effusion   11.87%
Emphysema   2.24%
Fibrosis   1.50%
Hernia   0.20%
Infiltration   17.74%
Mass   5.15%
No Finding   53.83%
Nodule   5.64%
Pleural_Thickening   3.01%
Pneumonia   1.27%
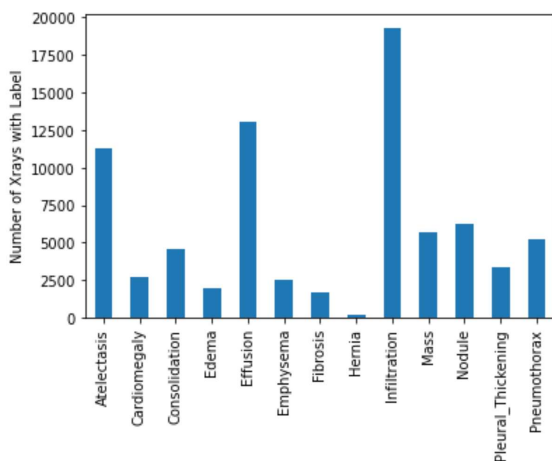Pneumothorax   4.72%

These findings can co-occur with pneumonia, the top 10 most occurring situations that involve pneumonia in the dataset are the following:
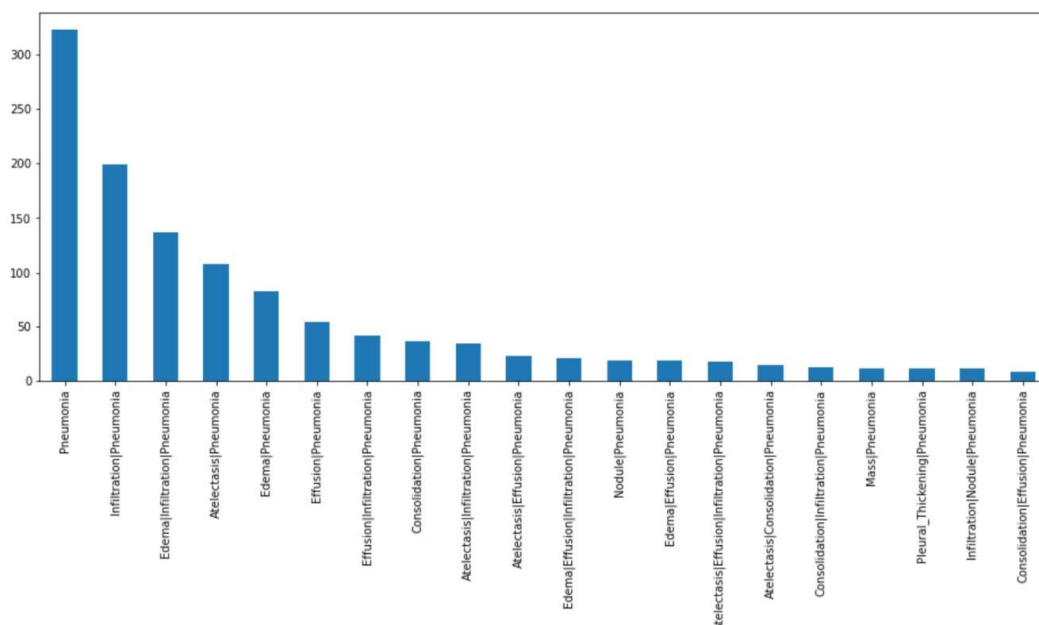
Label by label, the prevalence of the diseases among pneumonia cases are as follows:
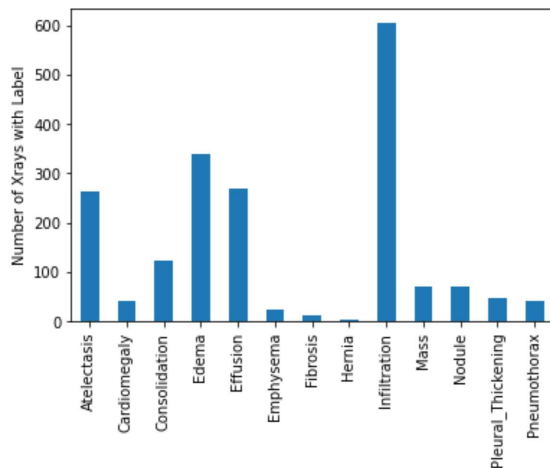

Label by label, the prevalence of the diseases among non-pneumonia cases are as follows:

The training set represents 2290 images randomly chosen from the initial dataset with the condition that half of the cases presents the label pneumonia and half of them does not. The average age is 45.64 years (STD=17.32). The male, female proportions are 59.26% and 40.74%. The 'PA', 'AP' position views proportions are 48.03% and 51.97%:

Description of Validation Dataset: The validation dataset also comes from the chest xray NIH dataset and is distinct of the training dataset Validation set should be as close as possible to 'reality' (here the initial dataset: 1.27% of pneumonia case). This value is too small to allow a good validation process, so the validation set was built upon the lasting cases (after training set construction) such as the proportion pneumonia:non-pneumonia cases is 1:3. The average age is 46.73 years (STD=16.69). The male, female proportions are 56.21% and 43.79%. The 'PA', 'AP' position views proportions are 43.09% and 56.91%:
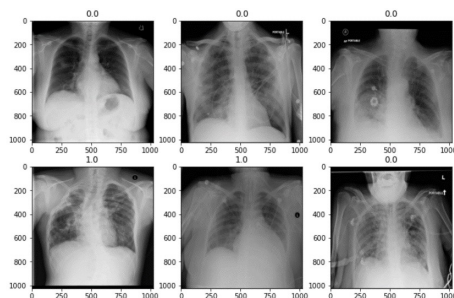
5. Ground Truth

The ground truth was the labels presented with the dataset on kaggle obtained by Wang and al. [arXiv:1705.02315]. The labels considered as ground truth were acquired through NLP algorithm and the accuracy is estimated being at least 90% accuracy [arXiv:1705.02315].

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: the validation dataset has been obtained from 4 clinical partners and consists of 120000 DICOM files of chest xrays : embedded image modality are systematically 'DX'. The imbedded DICOM xrays images are half 'PA' and half 'AP' view position of the chest with body part examined labeled 'CHEST', 'RIBCAGE', 'LUNG' and 'CHESTABDOMEN'. DICOM Image photometric Interpretation are 'MONOCHROME1' or 'MONOCHROME2'.

The age range of the patient goes from 1 year to 99 years with an average age of 38.2 years. The proportions of male and female cases are 49% and 51% which is representative of the US population.

25% of the validation dataset present pneumonia but the dataset also includes 13 other diseases : 'Atelectasis', 'Cardiomegaly', 'Consolidation', 'Edema', 'Effusion', 'Emphysema', 'Fibrosis', 'Hernia', 'Infiltration', 'Mass', 'Nodule', 'Pleural_Thickening', 'Pneumothorax' . This diseases are present among both pneumonia and non-pneumonia cases.

Ground Truth Acquisition Methodology: the ground truth/gold standard is obtained via the 'Study Description' part of the DICOM files as given by the radiologists.

Algorithm Performance Standard: The standard performance is the mean of three radiologists f1 score as obtained in the study: [arXiv:1705.02315].