Pneumonia Detector from Chest X-Rays

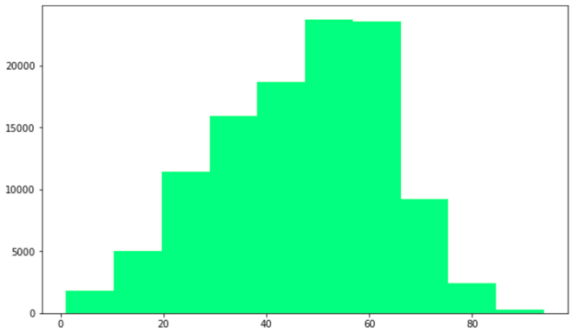
FDA Submission
Rafael Castro

Algorithm Description

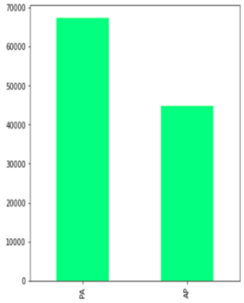
1. General Information

Intended Use Statement
The intended use for this algorithm is for patients from ages from 1 - 100 (NIH sample data was properly cleaned dropping outliers/errors from patients with age over 94+ for distribution visualization purposes only. However, training data is considering all samples and augmented images) who have been screened with Chest X-ray.

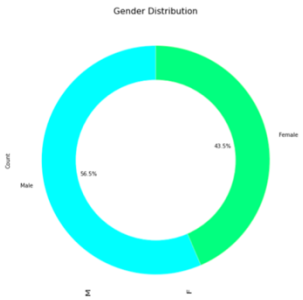
AGE DISTRIBUTION



X-RAY POSITION



GENDER MIX



Our algorithm was trained with data of 57% Males and 43% Females with a mean range of 47 - 65 years old, with 'PA' as predominant X-Ray position.

Indications for Use
Use as an additional resource to collaborate in prioritization and early detection of patients with suspects of pneumonia. Main users: radiologist, specialist, and first care.

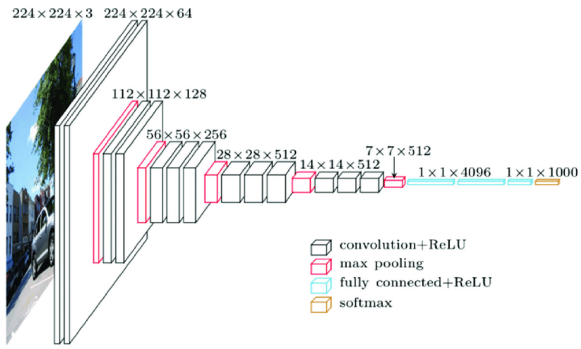
Device Limitations
Due to the presence of different other classes of disease (Effusion, Cardiomegaly, Mass, Atelectasis, Nodule, Hernia, Emphysema, etc.) may lead to false-positives pneumonia cases. The data only have posterior or anterior-posterior view X-Rays positions. Also, no prevalence history has been shared. However, evidence of pneumonia can be accurately detected by this algorithm and can be very helpful in order to raise a flag for the radiologist/other clinicians.

Clinical Impact Performance
Once the radiologist/clinician is warned, could effectively manage prioritizations in an automated way and all benefits of an early detection time-wise in all senses (patient- doctor).

Pneumonia Detector from Chest X-Rays

2. Algorithm Design and Function

We develop an algorithm that can detect pneumonia from chest X-rays helping practicing radiologists to diagnose showing evidence of presence or absence of pneumonia with a classification class model. Our algorithm, uses **VGG16 model** as a primary base.




```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
model_2 (Model)              (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_2 (Flatten)          (None, 25088)             0
_________________________________________________________________
dense_4 (Dense)              (None, 1024)              25691136
_________________________________________________________________
dropout_3 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_5 (Dense)              (None, 512)               524800
_________________________________________________________________
dropout_4 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_6 (Dense)              (None, 1)                 513
=================================================================
Total params: 40,931,137
Trainable params: 28,576,257
Non-trainable params: 12,354,880
_________________________________________________________________
```

The VGG network architecture was introduced by Simonyan and Zisserman in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition. This network is characterized by its simplicity, using only 3×3 convolutional layers stacked on top of each other in increasing depth. For our algorithm, other layers were added to optimize performance with dataset.

DICOM Checking Steps

Chest X-rays are currently the best available method for diagnosing pneumonia (WHO, 2001), playing a crucial role in clinical care (Franquet, 2001) and epidemiological studies (Cherian et al., 2005). In that regard, DICOM files have important information related to the patient including the pixel representation of the image. We are using python 'pydicom' library to read the files. First with the overall reading and then accessing the pixel representation.

Fig.1: DICOM reading:

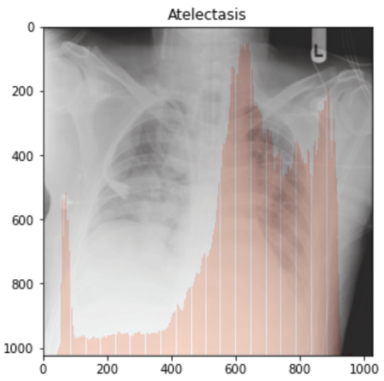
```
In [29]: no_finding_dcm = pydicom.dcmread('test1.dcm')

In [30]: no_finding_dcm
Out[30]: (0008, 0016) SOP Class UID                       UI: Secondary Capture Image Storage
         (0008, 0018) SOP Instance UID                    UI: 1.3.6.1.4.1.11129.5.5.11050364559275649246316982105025258226788
         (0008, 0060) Modality                            CS: 'DX'
         (0008, 1030) Study Description                   LO: 'No Finding'
         (0010, 0020) Patient ID                          LO: '2'
         (0010, 0040) Patient's Sex                       CS: 'M'
         (0010, 1010) Patient's Age                       AS: '81'
         (0018, 0015) Body Part Examined                  CS: 'CHEST'
         (0018, 5100) Patient Position                    CS: 'PA'
         (0020, 000d) Study Instance UID                  UI: 1.3.6.1.4.1.11129.5.5.11250701080328447820752201683219186696470 8
         (0020, 000e) Series Instance UID                 UI: 1.3.6.1.4.1.11129.5.5.11263085036218246837244082875521829335232 9
         (0028, 0002) Samples per Pixel                   US: 1
         (0028, 0004) Photometric Interpretation          CS: 'MONOCHROME2'
         (0028, 0010) Rows                                US: 1024
         (0028, 0011) Columns                             US: 1024
         (0028, 0100) Bits Allocated                      US: 8
         (0028, 0101) Bits Stored                         US: 8
         (0028, 0102) High Bit                            US: 7
         (0028, 0103) Pixel Representation                US: 0
         (7fe0, 0010) Pixel Data                          OW: Array of 1048576 elements
```

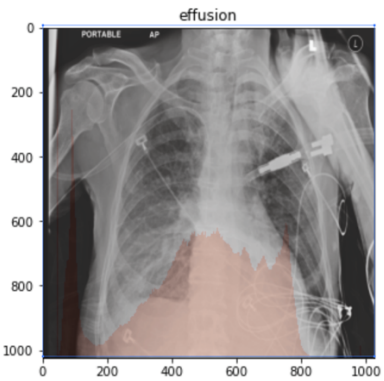
Pneumonia Detector from Chest X-Rays

Pre-processing steps
Understanding Image composition and distribution is key to determine zones of interest in which the algorithm is going to interact and identify binary class.

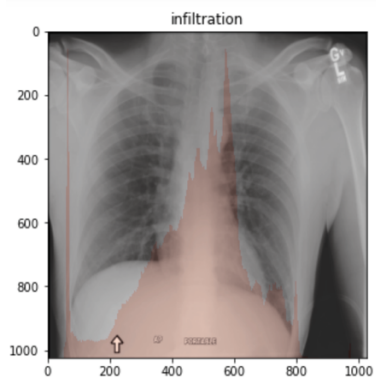
Intensity Distributions per other co-morbid diseases vs Pneumonia and No Pneumonia 🔍



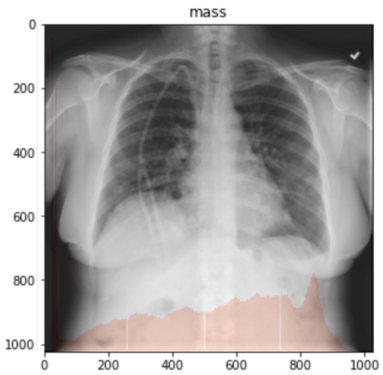
(a) Patient with **Atelectasis**
Mean = 156.67
Std = 58.45



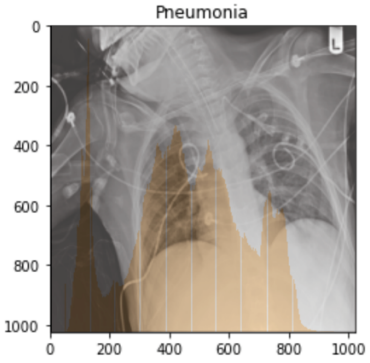
(b) Patient with **Effusion**
Mean = 157.86
Std = 49.64



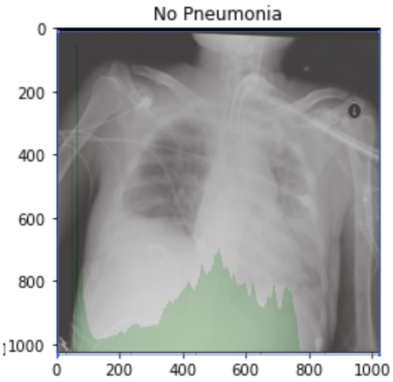
(c) Patient with **Infiltration**
Mean = 111.57
Std = 61.27



(a) Patient with **Mass**
Mean = 126.63
Std = 68.73



(b) Patient with **Pneumonia**
Mean = **107.05**
Std = **52.50**



(c) Patient with **No Pneumonia**
Mean = **108.23**
Std = **56.26**

As we stated at the Device Limitations part, the presence of other diseases may lead to false positives of pneumonia (for instance Atelectasis and Effusion shares a similar high mean, and particularly Effusion shares a low Std with Pneumonia, that could be also a co-morbid disease together with Pneumonia) in an optimized dataset, ***ideally the absence of prevalence nor other co-morbid diseases*** would be desirable. However, based on evidence regarding the histograms and considering the mean and std intensity of each image distributions, we can state that we can use de dataset to train the model, since tends to clearly mark the difference between a evidence or absence of pneumonia. With a little improvement to the algorithm, other diseases could be also identified.

Pneumonia Detector from Chest X-Rays

3. Algorithm Training

Parameters

The weights of the network are initialized with weights from a model pretrained on ImageNet (Deng et al., 2009). The network is trained end-to-end using Adam optimizer and using currently the largest publicly available chest Xray dataset, containing over 100,000 frontalview X-ray images with 14 diseases.

Types of augmentation used during training

Before inputting the images into the network, we downscale the images to 224×224 and normalize based on the mean and standard deviation of images in the ImageNet training set. We also augment the training data with random horizontal flipping

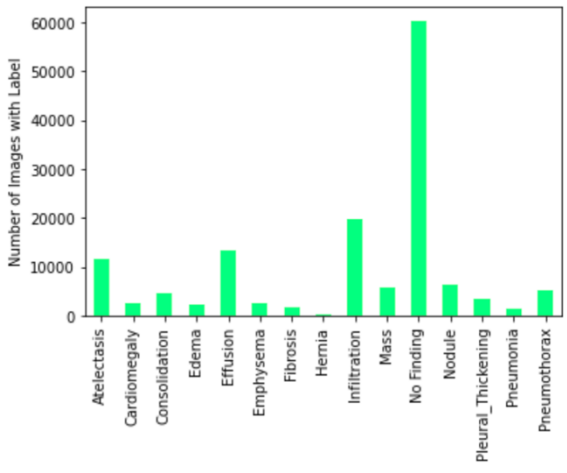
The rescale operation represents image reduction or magnification during the augmentation process. The rotation range denotes the range in which the images were randomly rotated during training, i.e., 5 degrees. Width shift is the horizontal translation of the images by 0.1 percent, and height shift is the vertical translation of the images by 0.05 percent. In addition, a shear range of 0.1 percent clips the image angles in a counterclockwise direction. The zoom range randomly zooms the images to the ratio of 0.15 percent, and finally, the images were flipped horizontally.

4. Databases

We are using NIH Database, this data includes 14 pathologies labels

```
In [12]: all_xray_df[all_labels].sum()/len(all_xray_df)

Out[12]: Atelectasis          0.103102
         Cardiomegaly         0.024763
         Consolidation        0.041631
         Edema                0.020535
         Effusion             0.118784
         Emphysema            0.022444
         Fibrosis             0.015040
         Hernia               0.002025
         Infiltration         0.177435
         Mass                 0.051551
         No Finding           0.538362
         Nodule               0.056475
         Pleural_Thickening   0.030187
         Pneumonia            0.012756
         Pneumothorax         0.047287
         dtype: float64
```

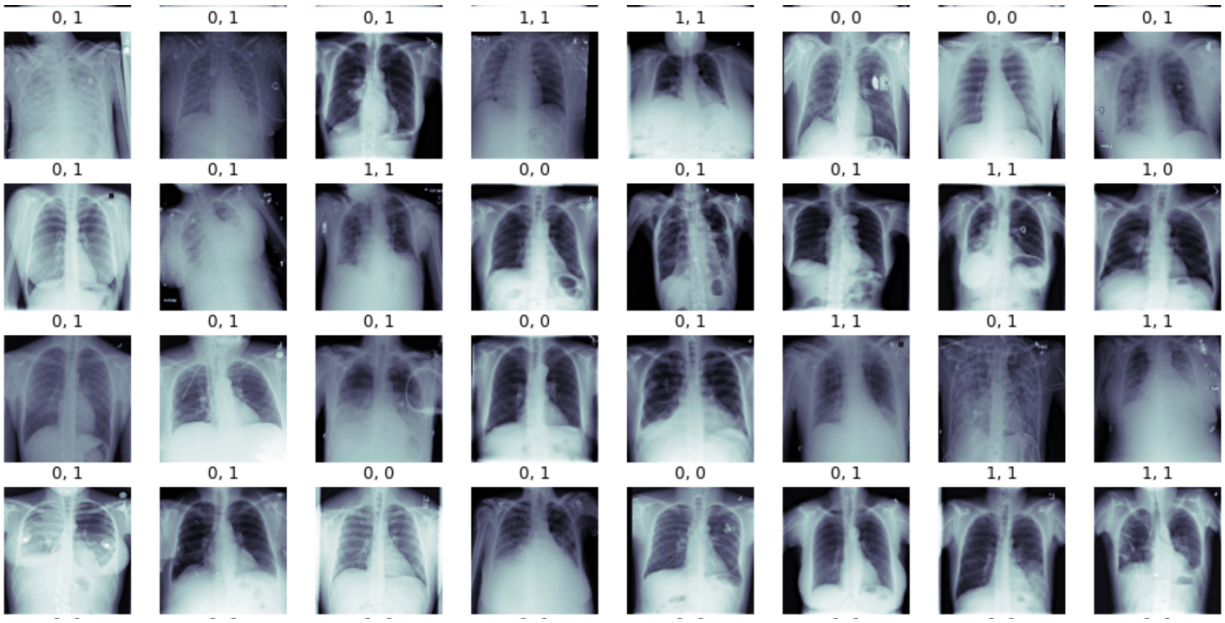


Pneumonia Detector from Chest X-Rays

Algorithm Training Visualization

We developed a model to detect and classify pneumonia from chest X-ray images taken from frontal views at high validation accuracy. The algorithm begins by transforming chest X-ray images into sizes smaller than the original. The next step involves the identification and classification of images by the convolutional neural network framework, which extracts features from the images and classifies them. Due to the effectiveness of the trained CNN model for identifying pneumonia from chest X-ray images, the validation accuracy of our model was significantly higher when compared with other approaches. To affirm the performance of the model, we repeated the training process of the model several times, each time obtaining the same results. To validate the performance of the trained model on different chest X-ray image sizes, we varied the sizes of the training and validation dataset and still obtained relatively similar results.

Fig. 3 Example predicted images with our algorithm (with our chosen **threshold of 0.49**) vs truth:



5. Ground Truth

Provided Ground Truth is ***Golden Ground*** truth, since the data includes one label classifying pneumonia positives plus other co-morbid diseases together.

Pneumonia Detector from Chest X-Rays

6. FDA Validation Plan

Patient Population Description for FDA Validation Data Set

Ideal dataset would be collected from chest X-Rays for the ***population range 1-100*** in similar ***gender proportion***. ***No prevalence of Pneumonia nor other diseases*** as Effusion, Nodules, Cardiomegaly, Atelectasis, Consolidation, Edema, Ephysema, Fibrosis, Hernia, Mass, etc., is required. A 80-20 Data Set Split (50-50 balanced), conserving real-life prevalence is also mandatory. Error in labeling might be a limitation if occurred.

The ***imaging modalities*** used in biology and medicine are based on a variety of energy sources, including light, electrons, lasers, X-rays, radionuclides, ultrasound and nuclear magnetic resonance. The images produced span orders of magnitude in scale, ranging from molecules and cells to organ systems and the full body (Rob, Richard A., 2009). Our algorithm uses ***Chest X-Rays*** from PA and AP position views to train on, ***imaging body parts*** such as, lungs, heart and bones, to understand for instance, a co-morbid disease such as Cardiomegaly or a Mass.

Pneumonia accounts for a significant proportion of patient morbidity and mortality (Gon¸calves-Pereira et al., 2013). Early diagnosis and treatment of pneumonia is critical to preventing complications including death (Aydogdu et al., 2010). With approximately 2 billion procedures per year, chest X-rays are the most common imaging examination tool used in practice, critical for screening, diagnosis, and management of a variety of diseases including pneumonia (Raoof et al., 2012). However, two thirds of the global population lacks access to radiology diagnostics, according to an estimate by the World Health Organization (Mollura et al., 2010). There is a shortage of experts who can interpret X-rays, even when imaging equipment is available, leading to increased mortality from treatable diseases (Kesselman et al., 2016).

Ground Truth Acquisition Methodology

We define **Silver Standard** as our ***optimal ground truth acquisition methodology*** since detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists. In that regard, the ideal ground truth would be weighting the radiologists labeling according to their experience and compare with the results of our algorithm.

Pneumonia Detector from Chest X-Rays

Algorithm Performance Standard

Pathology	Wang et al. (2017)	Yao et al. (2017)	Algorithm (ours)
Pneumonia	0.633	0.713	**0.62**

Fig. 4 Performance metrics description: AUC, Precision-Recall Curve, Score vs Threshold

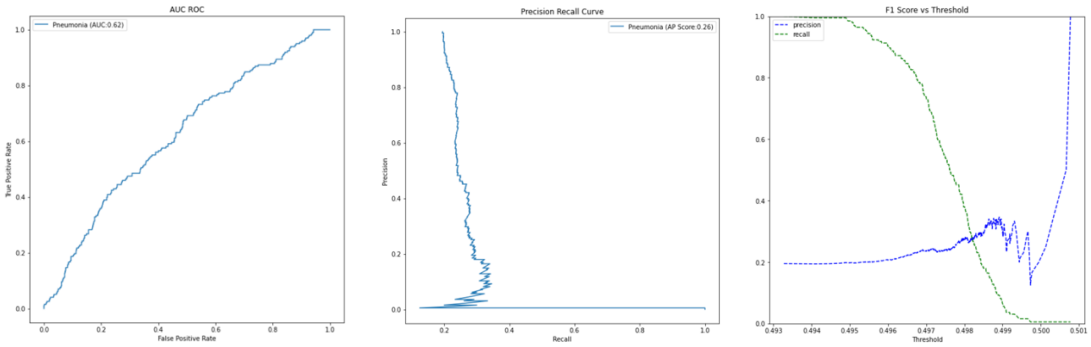


Fig. 5 Performance Metrics Results and benchmark

Performance Metric	Radiologist Avg.	Algorithm (ours)
F1 Score	0.387	**0.374**

```
In [47]: # Look at the threshold where recall is 0.8
         recall_value = 0.8
         idx = (np.abs(recall - recall_value)).argmin()
         print('Precision is: '+ str(precision[idx]))
         print('Recall is: '+ str(recall[idx]))
         print('Threshold is: '+ str(thresholds[idx]))
         print('F1 Score is: ' + str(calc_f1(precision[idx],recall[idx])))

         Precision is: 0.24421965317919075
         Recall is: 0.8009478672985783
         Threshold is: 0.49668416
         F1 Score is: 0.3743078626799557
```

References

Pranav Rajpurkar * 1 Jeremy Irvin * 1 Kaylie Zhu 1 Brandon Yang 1 Hershel Mehta 1 Tony Duan 1 Daisy Ding 1 Aarti Bagul 1 Robyn L. Ball 2 Curtis Langlotz 3 Katie Shpanskaya 3 Matthew P. Lungren 3 Andrew Y. Ng CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. URL https://arxiv.org/pdf/1711.05225.pdf

Hindawi Journal of Healthcare Engineering Volume 2019, Article ID 4180949, 7 pages https://doi.org/10.1155/2019/4180949