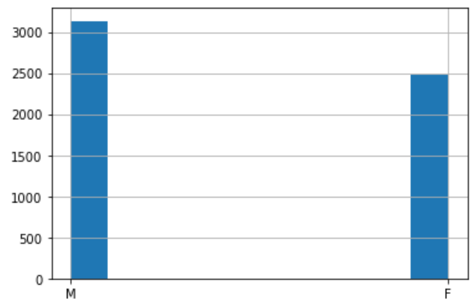
Nini Desmond

Pneumonia Detector

1. General Information



The frequency of males in our dataset tops that of females significantly. This would be of relative advantage to the male gender features which we shall use in feeding our model

INTRODUCTION:

According to Wikipedia.org, each year, pneumonia affects about 450 million people globally (7% of the population) and results in about 4 million deaths.

Chest X-rays are currently the best available method for diagnosing pneumonia (WHO), playing a crucial role in clinical care.

However, detecting pneumonia in chest X-rays is a challenging task that relies on

the availability of expert radiologists. In this work, we present a model that can automatically detect pneumonia from chest X-rays at a level exceeding practicing

radiologists.

Intended Use Statement:

This algorithm is intended to use for Pneumonia patients who have been administered to a chest X-ray screening and had never demonstrated a chest abnormalities.

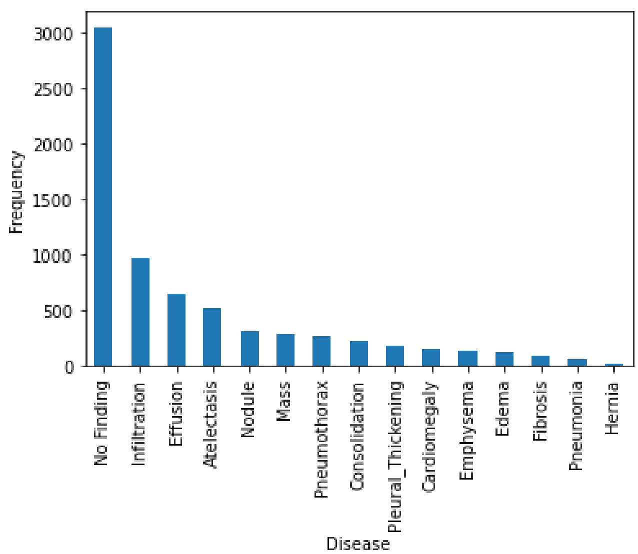
OBJECTIVE:

Indications for Use:

The aim of this study was to develop an algorithm that can evaluate the performance of the processing system in extracting pneumonia-related concept from chest X-ray reports.

Device Limitations and Clinical Impact of Performance:

This dataset has a limitation of targeted victims This would not be best for the society given that this disease does not discriminate in affecting potential patients. And looking at this graph below, we can see that the dataset comprises of just 14 diseases, of which only one is of interest to our algorithm.

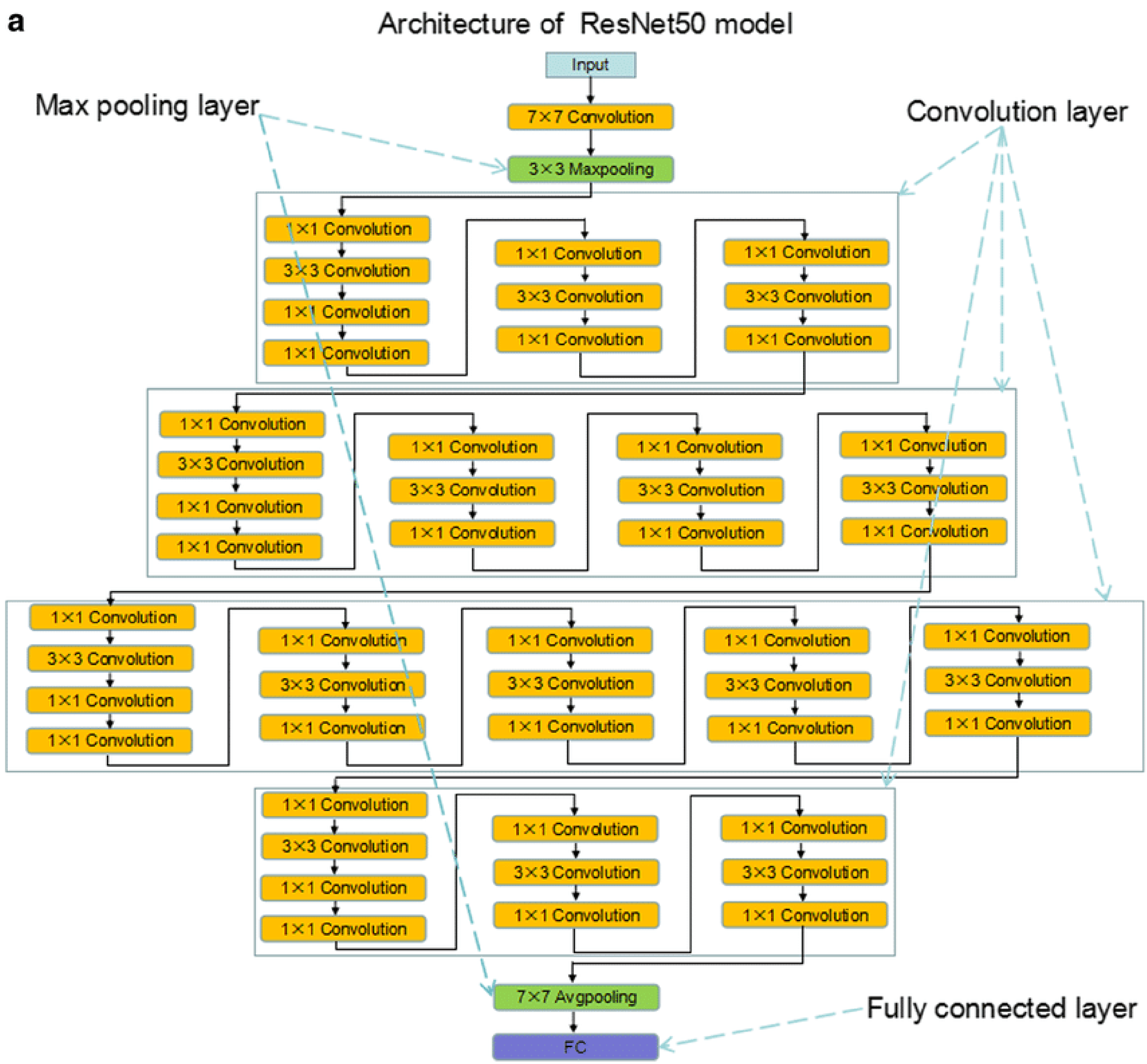


Because of the presence of these other diseases it will be a downside for us to record a false positive situation on a subject. This patient will be under the influence of a poor prediction, which may wreak havoc. And on the other hand, the model will actually help patience in automating the process of detecting Pneumonia from Chest X-Ray and saving them a lot of time for example.

Also, since we are more focused on *recall* instead of *precision*, If the algorithm say patient has pneumonia then it is possible that it may not have pneumonia.

2. Algorithm Design and Function

Capture depicts the architecture of ResNet50 and deep learning



A Resnet50 pre-trained model was used. One major reason for choosing this network was because of its ability to it avoids negative outcomes. If the network's depth is increased. And it accelerates the speed of training So we can increase the depth and still have fast training.

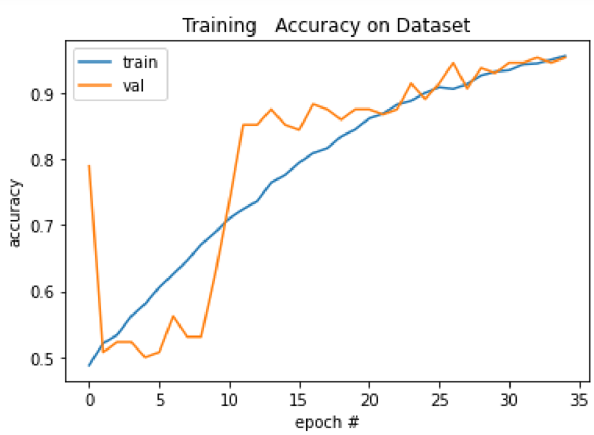
The checked **XRAY** DICOM file used was of the body part **chest** and with a view position of **PA** and **AP**

DATA PROCESSING:

This solution respectively split a dataset of 10000 images to 8000 by 2000 images of the train and validate. And to a class of 2 labels with a datatype of 0 and 1 integers. One represents positive cases with Pneumonia and Zero represents negative cases with other diseases. This would imply we have a binary problem to be treated in building our pre-trained model.

Also, rescaling was set to 1/255 as it to multiply our data before processing to easily target classes between 0 and 1.

To curb overfitting, **Data augmentation** was applied to our training set, which is just a strategy that enables practitioners to significantly increase the diversity of **data** available for training models, without actually collecting new **data**. Some of these techniques adopted were **Rescaling, Rotating, zooming**



This was the result of training up to 35 epoch.

ResNet-50 is a convolutional neural network that is 50 layers deep. One can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224

3. Algorithm Training

35 epochs were a good value from tweaking other parameters to avoid overfitting

A Stochastic gradient descent Optimizer was used with a learning of as low as 0.0001 which proved to be very effective in reducing the losses.

The **binary_crossentropy** Loss function was equally used as it smartly minimized the distance between two **Y_true and _Y_pred**

Eventually, a Dropout with probability 0.5 was added to the model to ensure the successful flow of training to the designated Epochs

A Batch size of 128

About the layers, first 24 weights were frozen and the following layers were added

```
            # add new classification head
    x = GlobalAveragePooling2D()(base_model.output)
    x = Dense(2, activation='sigmoid')(x)
    predictions = Dense(1, activation='sigmoid')(x)
    my_model = Model(inputs=base_model.input, outputs=predictions)
```

METHODS:

The pneumonia detection task is a binary classification problem, where the input is a frontal-view chest X-ray denoted as image, while the label is a binary classification. This is exactly what we have done in this submission to ensure that we have a clear case of prediction (pneumonia [1] or not pneumonia [0]).

Before inputting the images into the network, we downscale the images to 32×32. And a bunch of 128 single phase of batch size was chosen, which turned out to fit well for clustering predictions. A Resnet50 pre-trained model was used. One major reason for choosing this network was because of its ability to it avoids negative outcomes. If the network’s depth is increased. And it accelerates the speed of training So we can increase the depth and still have fast training.

Looking at the first DICOM image, the header was very informative, depicting a property of the image to be tested, which had a Study description field that holds the value of our labels column from our EDA sampled data frame. This tells us that we are definitely about to test an image that falls under a certain labeled class group (0). And of course the Patient's sex follows suit with a male label of age 80. This would mean we should not be expecting a positive case for our Pneumonia testing.

Final Threshold and Explanation:

So, since we are focused on the algorithm giving more of the relevant results, and we will have to avoid the false cases at all cost and confined our scope with the positive cases. This is where **Recall** comes in. We are going to take advantage of its computation over a PRECISION formula

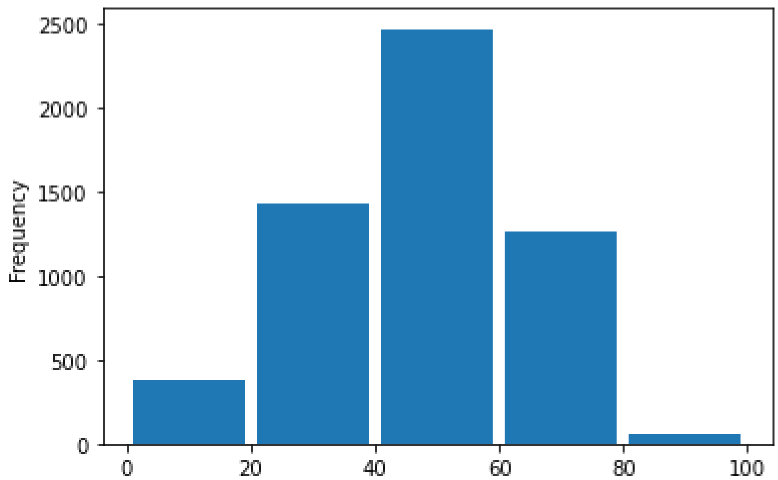
Below depicts the end result of this case study

```
Precision is: 0.018691588785046728
Recall is: 1.0
Threshold is: 0.41653234
F1 Score is: 0.03669724770642201

 -------------------- High recall --------------------

Precision is: 0.009433962264150943
Recall is: 0.5
Threshold is: 0.41694063
F1 Score is: 0.018518518518518517
```

4. Databases

This NIH Chest X-ray Dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. To create these labels, the authors used Natural Language Processing to text-mine disease classifications from the associated radiological reports.

The gender in slightly imbalanced as we could see as above, with more males than females. And this dataset is of age group between 5 to 90



The Dataset has been divided in training and validation set following the ratio of 80 to 20. And the training set was further divided to a ratio of 1 :1 for Pneumonia and non-Pneumonia cases While the Validation set was further divided to a ratio of 1 is to 3. The validation data set doesn't have the same image preprocessing of training. In validation set, we are only resizing the images, no other additional pre-processing has been applied.

5. Ground Truth

The NIH Chest X-ray Dataset highly imbalanced in relation with the patients with pneumonia. NLP-derived labels from the NIH are sub-optimal since they are more general than only the case of Pneumonia, one patient can have more than one disease similar to Pneumonia and many of the with more prevalence in the dataset than Pneumonia. This could impact the algorithm clinical performance. Also, the image labels were NLP-extracted so there could be some wrong labels which could render a poor result

A benefit would be that the NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community. Also, it can be used to reduce wasteful (and expensive and error prone) duplication of data

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset:

Ground Truth Acquisition Methodology:

Algorithm Performance Standard:

CONCLUSIONS:

The study of building X-ray models for pneumonia remains top-notch. It is quite interesting to see the outcome.

An ideal dataset should have relevant information (Pneumonia) cases for both males and females of all ages. The more the merrier. The ratio of Pneumonia cases to that of other diseases was not even proportional (322 to 111798). I think having the reciprocal of this will be a good choice for feeding our model for it to learn after splitting the data to a valid and train set. Images with many features like different genders have so many common features (characteristics of a male and female). Their postures, sizes, and maybe weight will definitely serve as unique features to help our model male the classes for prediction

An ideal Dataset would be the one that entails all relevant details like the current Dicom file information and even more of DX. Image type of "DX" and the only body part examined is the chest imaging modality, body parts examined, age range, gender distribution, prevalence of pneumonia (and other diseases). I think adding a Patient's occupation to the header can help in fostering prediction after labeling a class. This is important because it is very common for a laborer of a common group to contaminate diseases like Pneumonia as an industrial disease. This will definitely be a great feature to embed into our model

For ground truth, we need to label each of our data collected with class label whether the x-ray image is of pneumonia patient or Non-Pneumonia(Healthy) patient.

We recorded an f1 score of 0.35 which doesn't appear to be above an average score based on this paper

And a silver standard approach of using several radiologists would be optimal as seen here