Classify Chest X-ray Images

by Y. Kostrov

Contents

Overview

In this project, I create a model that will classify x-ray image of a chest as the one that has pneumonia or the one that doesn't have pneumonia

Business Problem

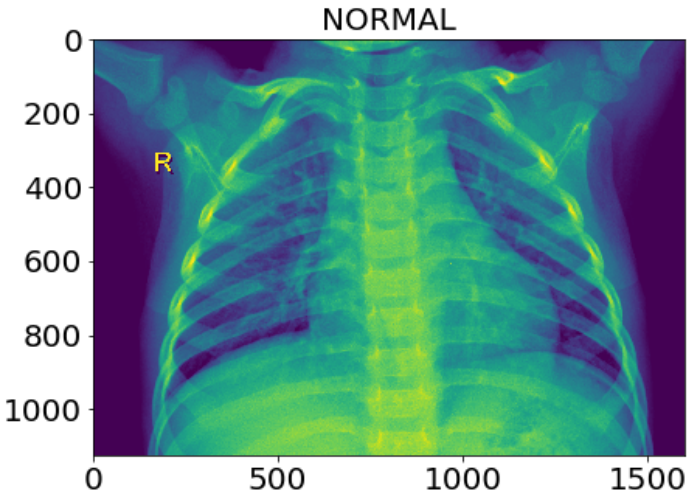
- Pneumonia is a very dangerous diseases that is caused by a bacterial or viral infection of the lungs.
- The consequences of pneumonia could be catastrophic within a short period of time if not diagnosed quickly.
- In this project I developed two models for the classification of chest X-ray images into NORMAL (no pneumonia) vs. PNEUMONIA(there is pneumonia in the lungs).

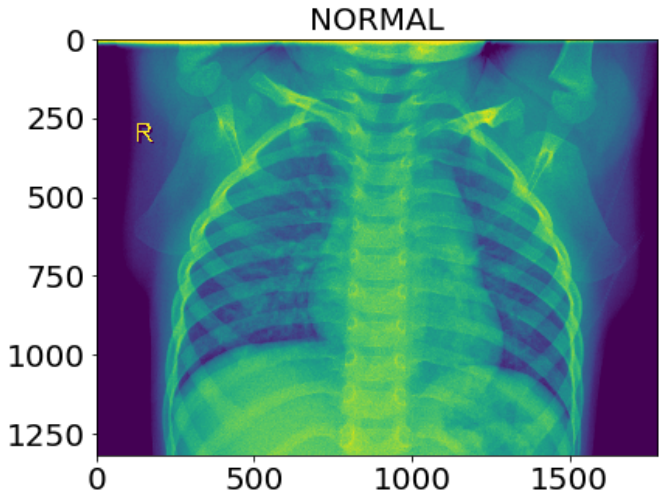
Data Used in the Project

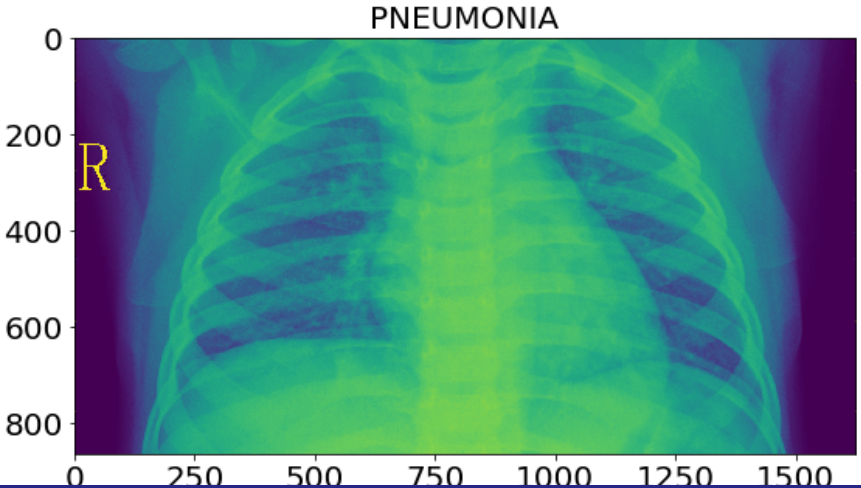
- Data for this project was downloaded from Mendeley Data: Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images.
- Data consists of chest x-ray images of different size split into two directories for training and testing.

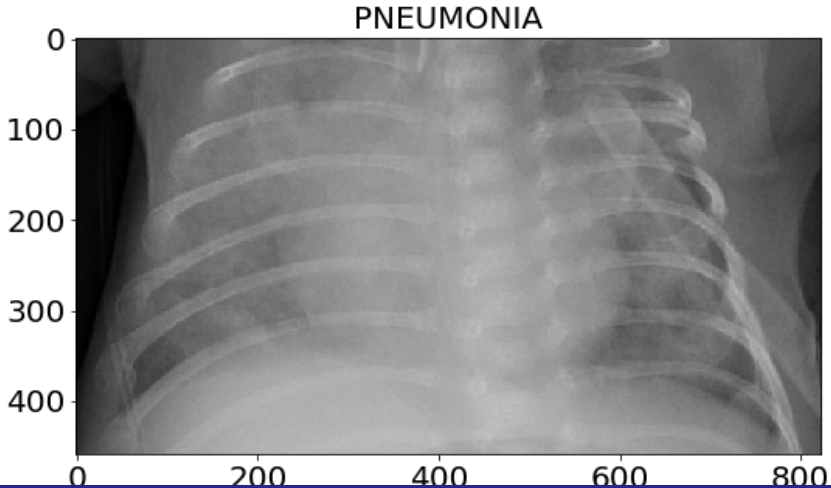
Number of Normal vs Pneumonia in Train Directory

Sample Images from Train Directory

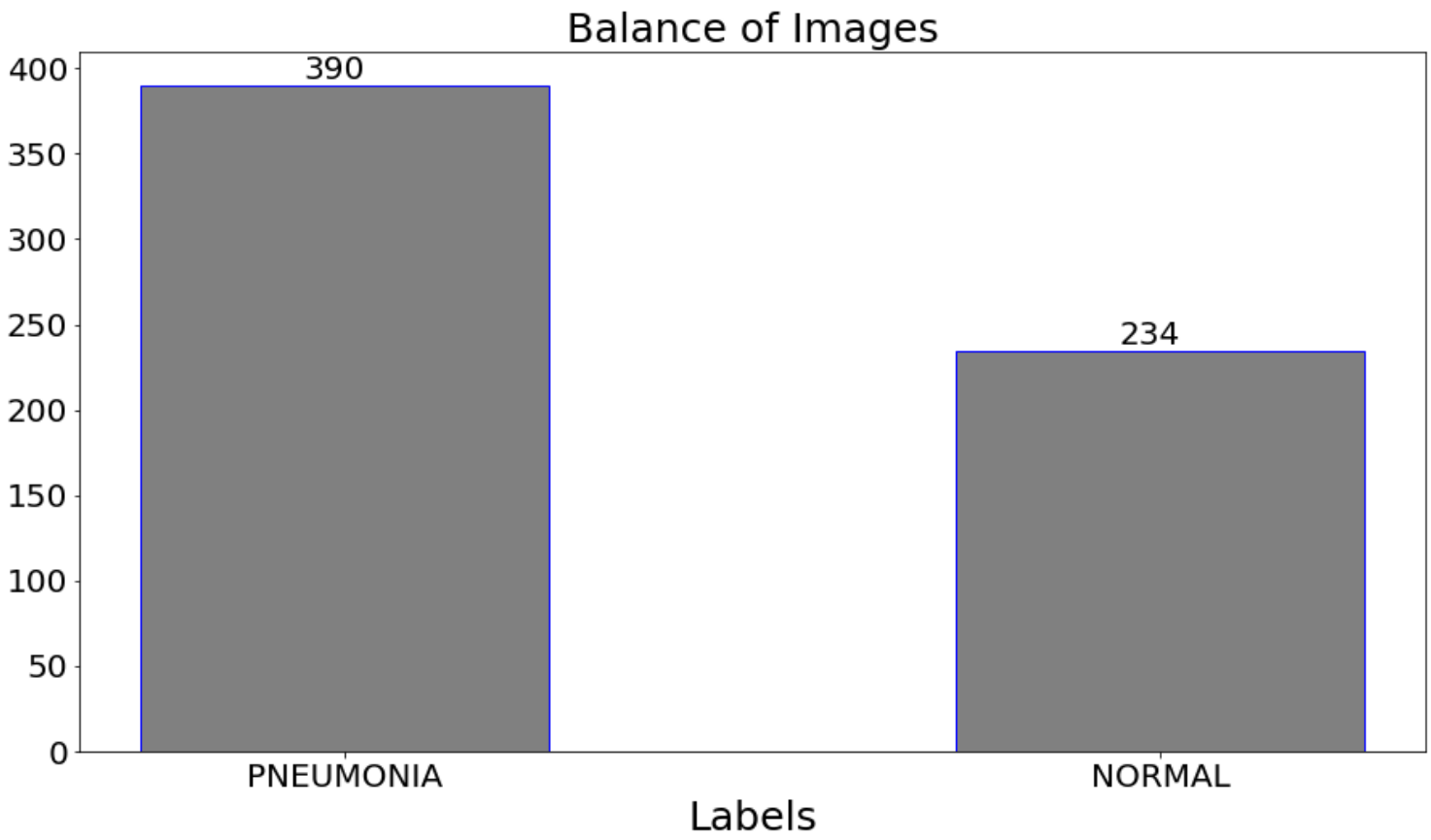




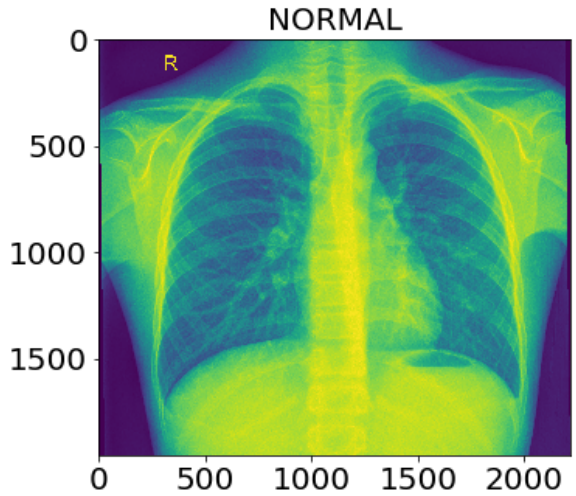


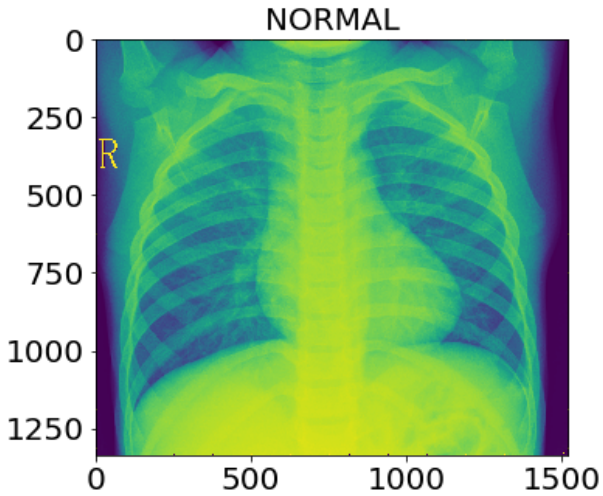


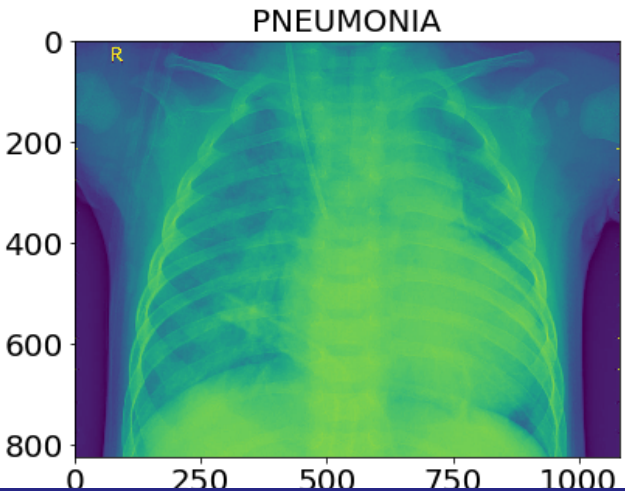
Number of Normal vs Pneumonia in Test Directory

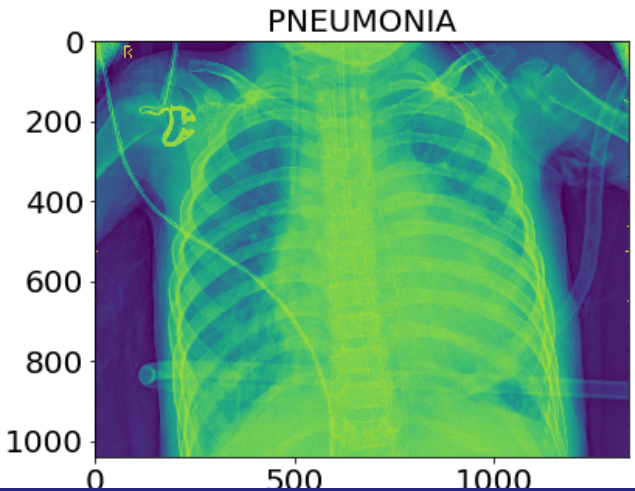


Sample Images from Test Directory









Modeling: Creating Models

I have built the following two models

- Baseline Model from scratch
- Transfer Learning Model based on VGG19 trained on imagenet dataset.

I used recall as a primary metric and accuracy as a secondary metric for assessment of models.

Explanation of Recall

- Recall is defined as:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{\text{True Positive}}{\text{Total Actual Positive}}$$

Explanation of Recall

- Recall calculates how many of the Actual Positives our model captures by marking it as Positive (True Positive).
- Thus Recall is a better model metric when there is a high cost associated with False Negative.
- In our case False Negative is predicting images as "NORMAL" when in reality the image has the label "PNEUMONIA".
- In this case pneumonia will not be diagnosed in the patient and consequnces might be drastic.

Explanation of Recall

- For instance, in prediction of pneumonia.
- If the case of pneumonia is predicted as normal(Predicted Negative), then the person who is sick will not be treated on time.

Explanation of Accuracy

- There is another metric we will use is called "accuracy".
- Accuracy is defined as:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{All Predictions}}$$

Explanation of Accuracy

- Accuracy is the number of correctly predicted data points out of all the data points.
- Often, accuracy is used along with recall.

How Well Models Performed

- Baseline Model achieved 99% on the recall metric and 87.5% on accuracy score
- Transfer Learning Model achieved 97.4% on the accuracy metric and it is, also, has accuracy at 97.6%.

Business Suggestion

Based on my analysis,

- I suggest to use Transfer Learning model for the prediction of pneumonia in chest x-rays since it gets better balanced results between recall and accuracy metrics.

THE END
THANK YOU!