Chest X-Ray Classification

By Steven Jasper

Good morning. My name is Steven Jasper and I will be presenting the results of my classification algorithm regarding detection of pneumonia in chest x-rays.

Problem Statement

We need to increase accuracy for the detection of Pneumonia.

The problem that we are facing here is we need to increase the accuracy of detection for patients who may be suffering from pneumonia. This will allow us to provide life-saving medical treatment with less naked-eye analysis.

Business Value

- Decrease Time-to-Detection
- Minimize Risk of Human Error
- Provide Proof of Concept for Automated Illness Detection

The business value here is simple, we will decrease time-to-detection, reduce the risk of Human error, and provide a proof of concept for automated illness detection.

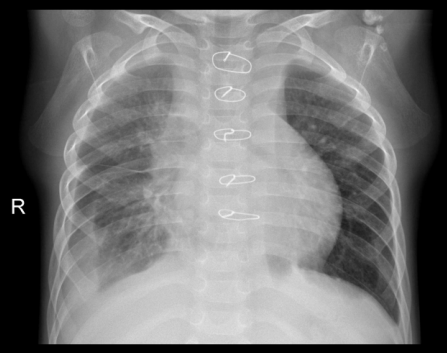
Methodology

OSEMN data science workflow, which involves:

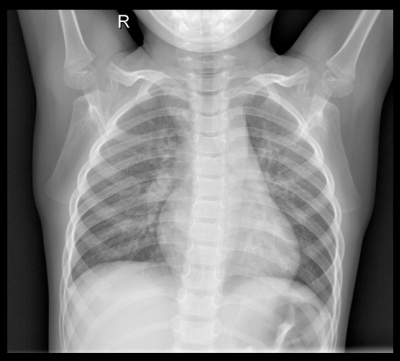
- Obtain (import the data)
- Scrub (clean the data, deal with missing values and data types)
- Explore (answer descriptive questions using EDA)
- Model (build our predictive model)
- iNterpret (comment on our model and findings)

I utilized the OSEMN data science work flow for this analysis. I began by obtaining the data, once the data was obtained I scrubbed, explored and began modeling our data. Finally we took outcomes from our model and interpreted these to produce our findings.

Findings



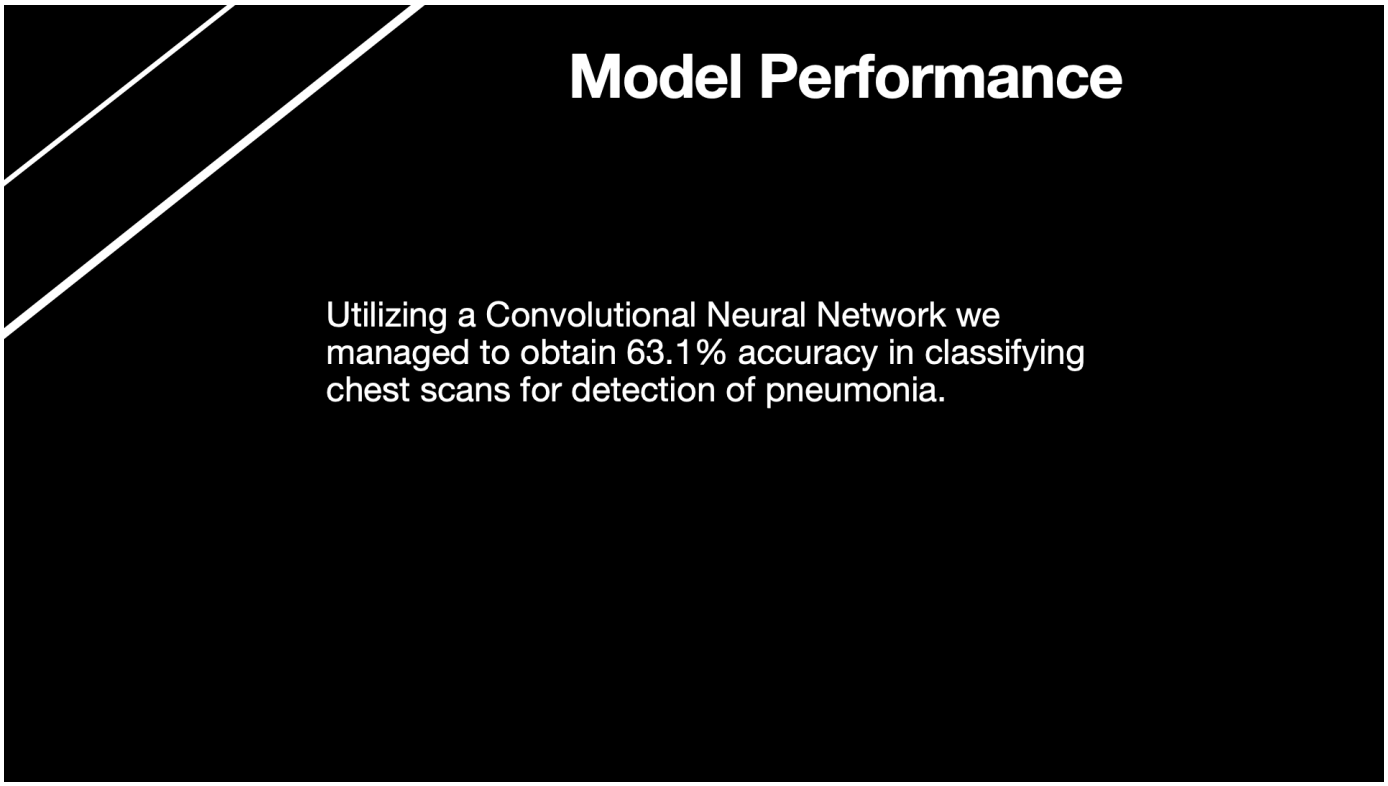
Pneumonia Positive



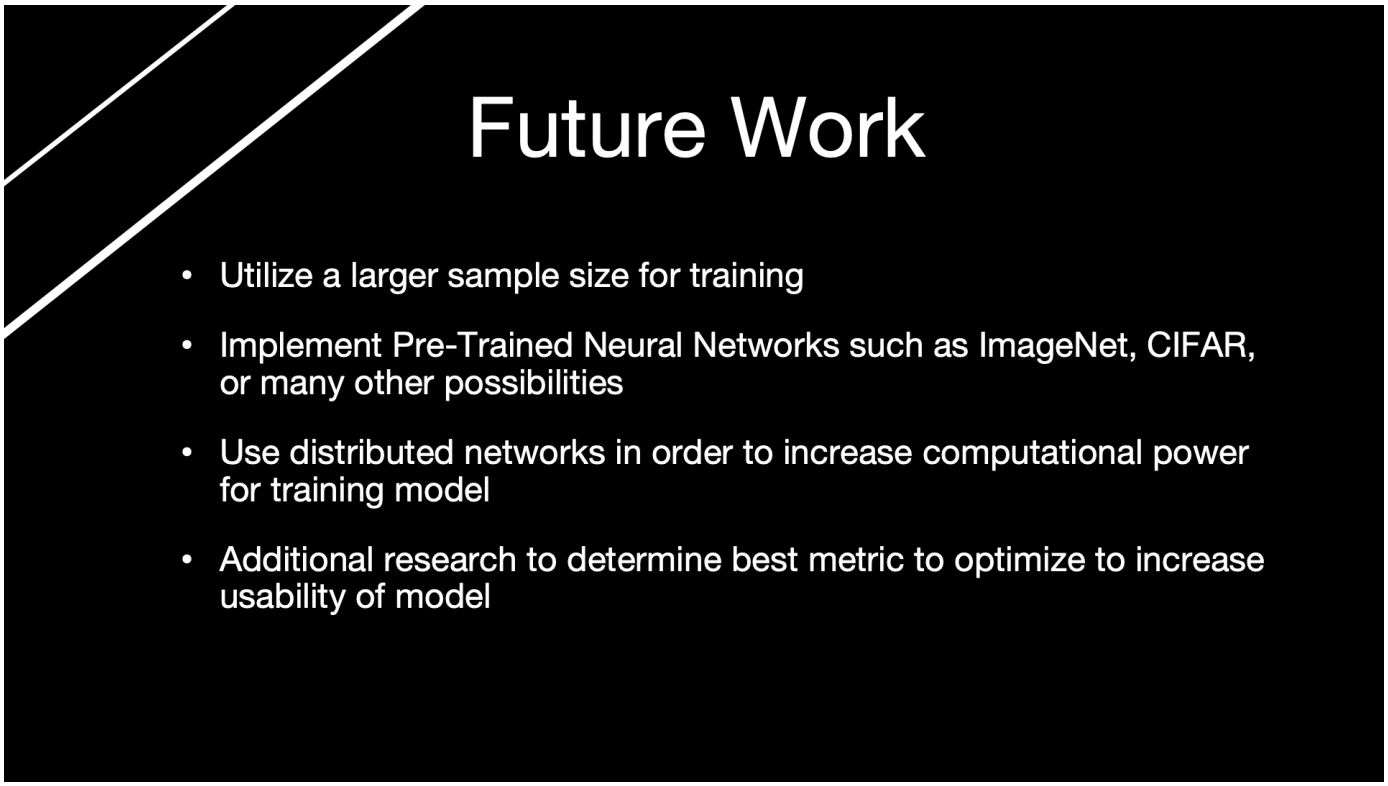
Pneumonia Negative

Here we take a look at two chest x-rays. At first glance it is incredibly difficult to determine which chest scan is positive for pneumonia, to the untrained eye. This leads to potentially ambiguous results even to some trained professionals, the idea is to reduce, and potentially remove, the human eye portion of this diagnosis. With images this close in detail the

chance for a False Positive or a False Negative become much higher, resulting in potentially more time-to-treatment ratio, which could also result in permanent damage to the patient’s lungs.



Utilizing our Convolutional Neural Network we obtained 63.1% accuracy with a relatively small data set, and limited computational resources. While the accuracy is relatively low, we do provide a proof of concept that with a larger sample size, more computational power, and if we expand our R&D time we can obtain a higher accuracy value for this network.



Some future work topics include utilizing a larger sample size, more data always increases model performance. I would also like to implement pretrained networks in order to increase layer performance of our CNN. In order to increase computational power in training our model we would want to explore distributed computation options. I also

would like to take a look at some research on the topic of the medical field to find the best metric of success for our model.

Questions?