

Pediatric Chest X-ray Classification for Pneumonia

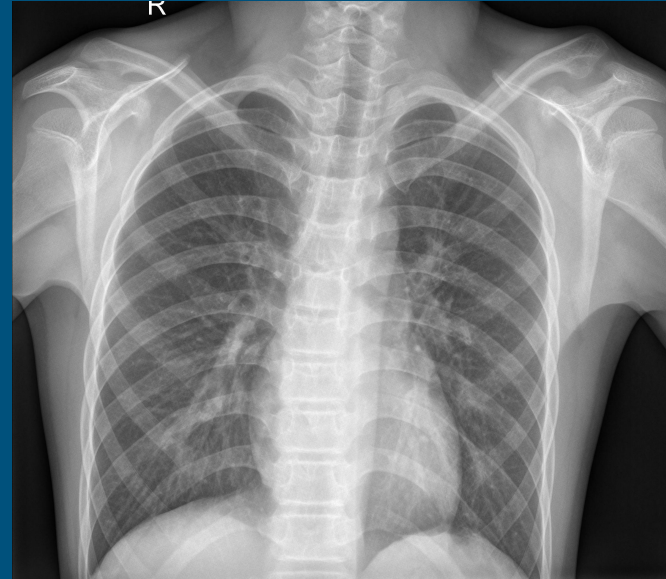
Anila Qureshi

Overview:

Methodology: [OSEMN]

- Binary Classification using a deep learning neural network called Convolutional Neural Networks: Used to predict Pneumonia
 - Obtain
 - Scrub
 - Explore
 - Model
 - iNterpret

PNEUMONIA V NORMAL X-RAY



Severity of the problem

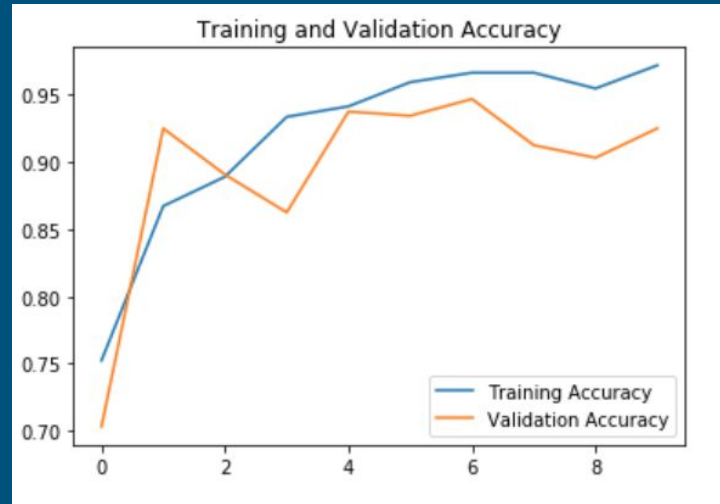
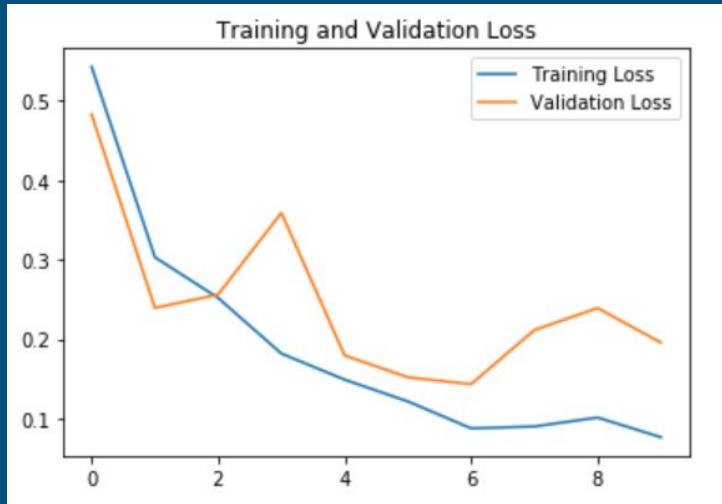
- Pneumonia is an infection of the lungs. Pneumonia accounts for 15% of deaths in children under the age of 5.
- It is caused by bacterial or viral infections, parasites or fungi.
- Pneumonia can be deadly and is preventable.
- Diagnostic image industry is a multibillion dollar industry dominated by the healthcare industry, specifically hospitals.
- Diagnostic errors cost the US economy about \$750 billion each year

Q1. What is the impact of data science inside medicine?

- Another way to help cut immeasurable costs as stated from the previous slide, many large hospitals around the world that dominate in research are opening data science departments (Yale is hiring! Also has one of the top medical programs in the world)
- Data science helps to aid in drug trials/discoveries, disease prevention, helps with diagnostics, treatment, post care monitoring, and human resource allocation

Baseline Model

- 3 Layers, with deep learning I have found from the lectures and now (little experience that when training a network we start off with less layers)

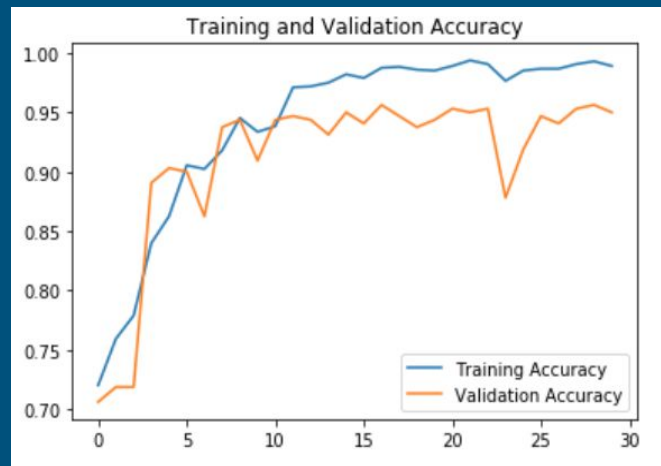
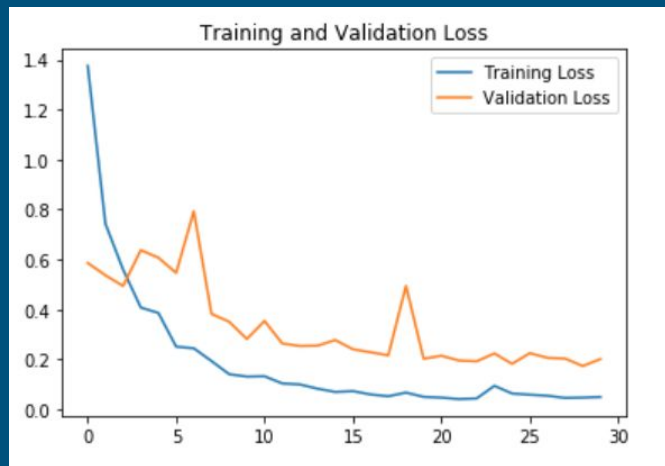


Q2. Should the healthcare industry focus on false positive or false negatives?

- False Negatives!
- Most often false positives are treated with antibiotics which are typically include azithromycin, penicillin, amoxicillin this does not harm the body long term unless there are severe allergies to penicillin
- Another thing to think about with False negatives are not just lack of treatment, but also misdiagnosis for other respiratory illnesses that could potentially be terminal.

Best Performing CNN Model

- 4 Layers, initially I had 6 but ran into some problems with gradient descent
- Dropout was added to ignore random units within the model to help reduce overfitting
- Also reduced nodes



Insight on Accuracy

- After running the winning model on to the test set we can see...
- Results_test = 1.7811826014518737, 0.7200000286102295] Not very accurate
- Winning model is with a threshold selection of >.3

```
array([[60,  1],  
       [26, 13]])
```

Q3. How can this problem be solved?

- More training, given the time of the contract work given, the ideal scenario would be to continue training these images with convolutional neural networks
- Gain more insights into the impacts of False Negatives, apply them to the winning model

Suggestions to Stakeholders

- Implement and train a convolutional neural network alongside each x-ray department to reduce the \$750 billion problem with image diagnostics
- second/third virtual opinions from India! India is slowly becoming the go to country for second medical opinions
 - Cost efficient
 - Quick