Pneumonia X-Ray Analysis

...

Bryan Valencia

Overview

- Taking a deeper look into the **image classification of x-rays** belonging to children between ages one and five and testing various models to be able to better identify patients who are suffering of pneumonia

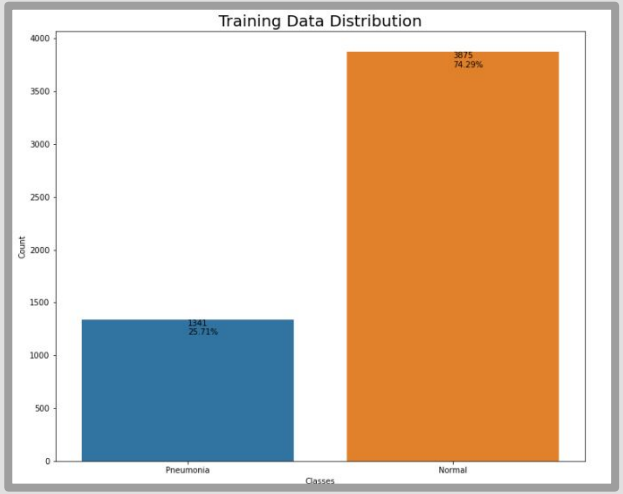
Business Problem

- Our stakeholders, Guangzhou Women and Children's Medical Center, are requesting for a model to be made that can be used to appropriately distinguish the difference between x-rays classified as "NORMAL" and "PNEUMONIA"
- With an efficient model at hand, the hospital can better **identify and combat** this sickness

Understanding the Data

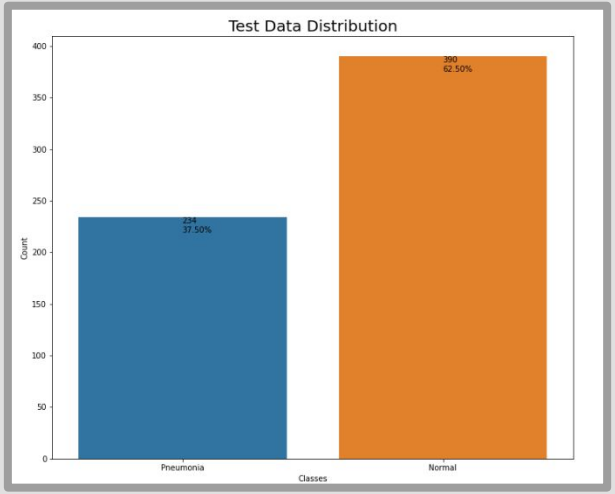
“NORMAL” vs “PNEUMONIA”

Training



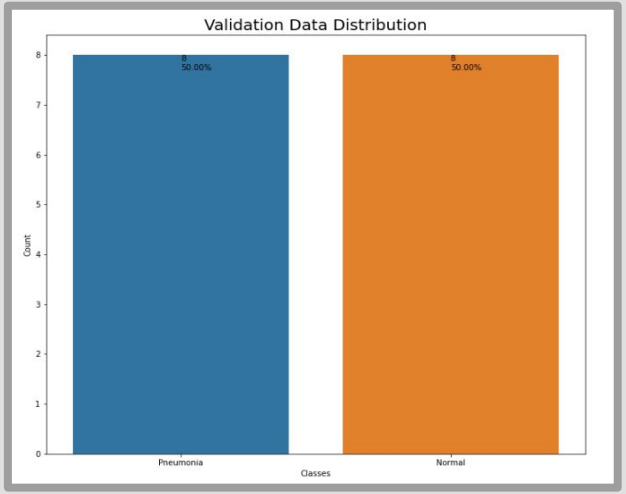
- Normal: 3,875
  - 74.29%
- Pneumonia: 1,341
  - 25.71%

Testing



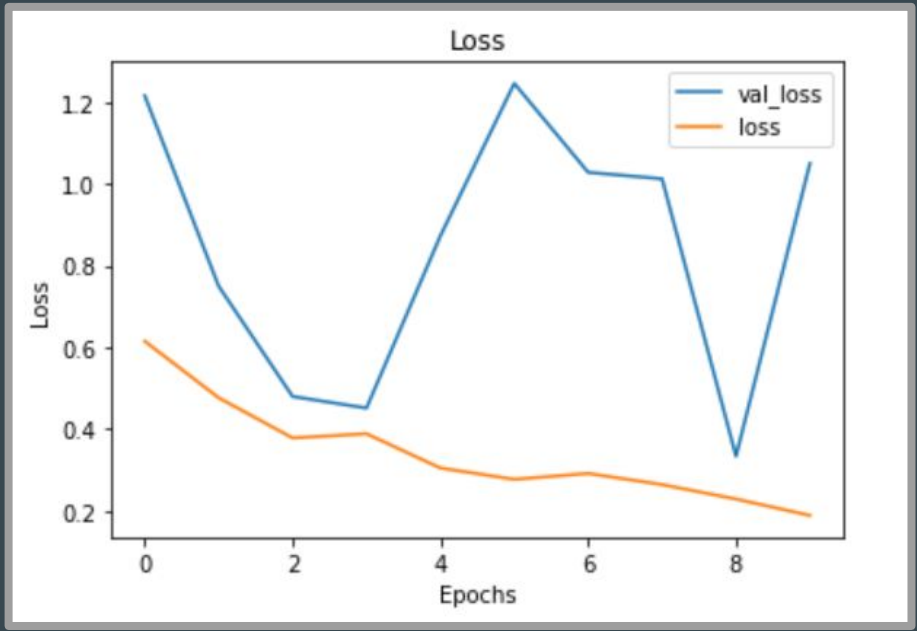
- Normal: 390
  - 62.50%
- Pneumonia: 234
  - 37.50%

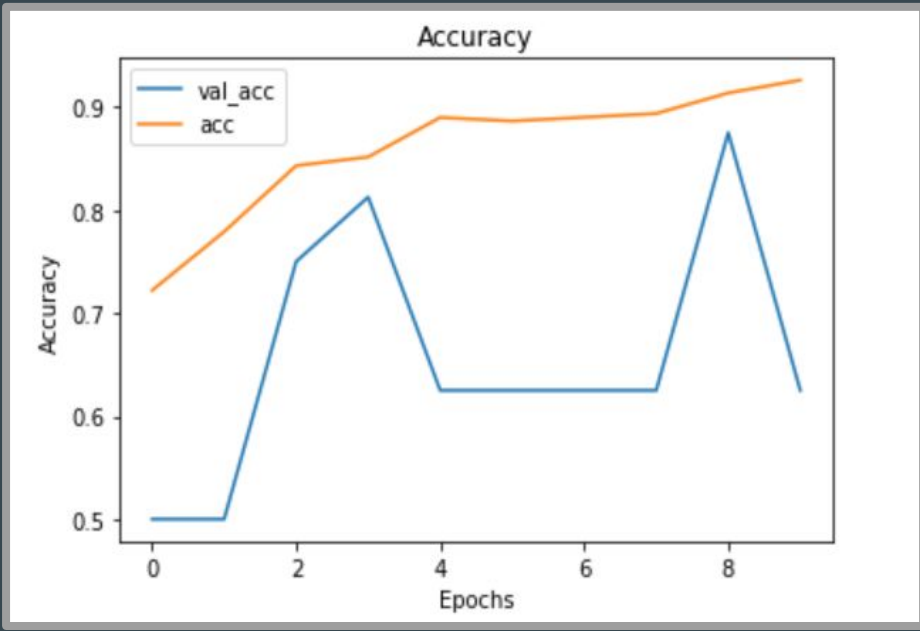
Validating



- Normal: 8
  - 50%
- Pneumonia: 8
  - 50%

Model 1 (Densely Connected Neural Network)

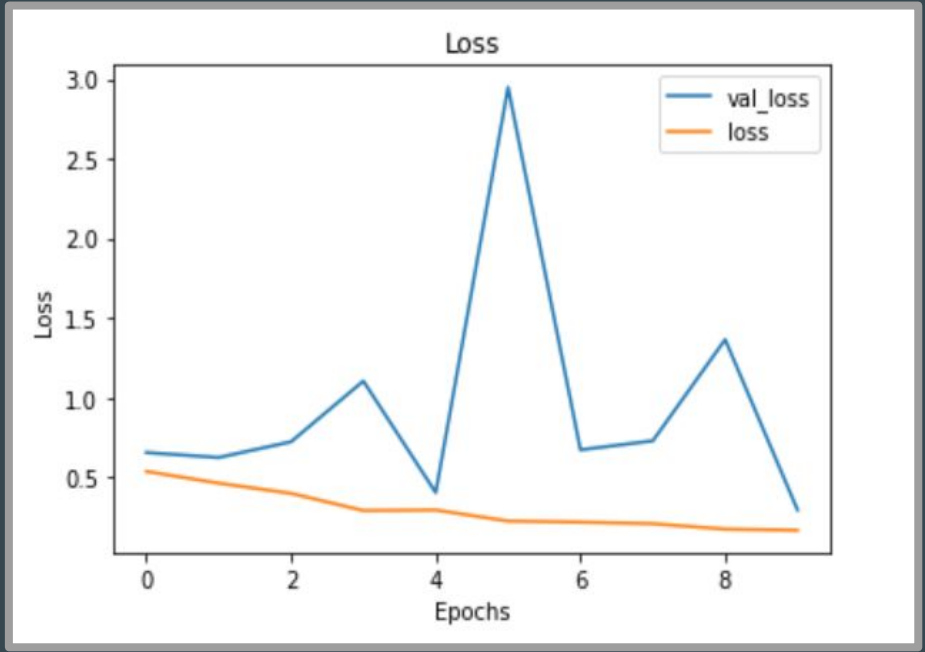


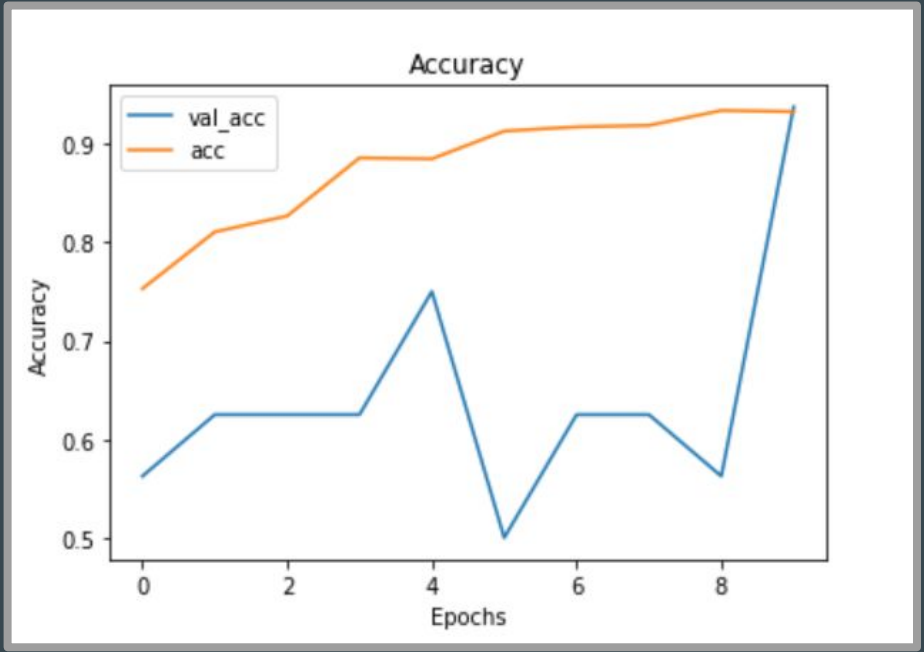


- Validation Loss: 1.04
- Training Loss: 0.28

- Validation Accuracy: 0.62
- Training Accuracy: 0.87

Model 2 (Convolutional Neural Network)

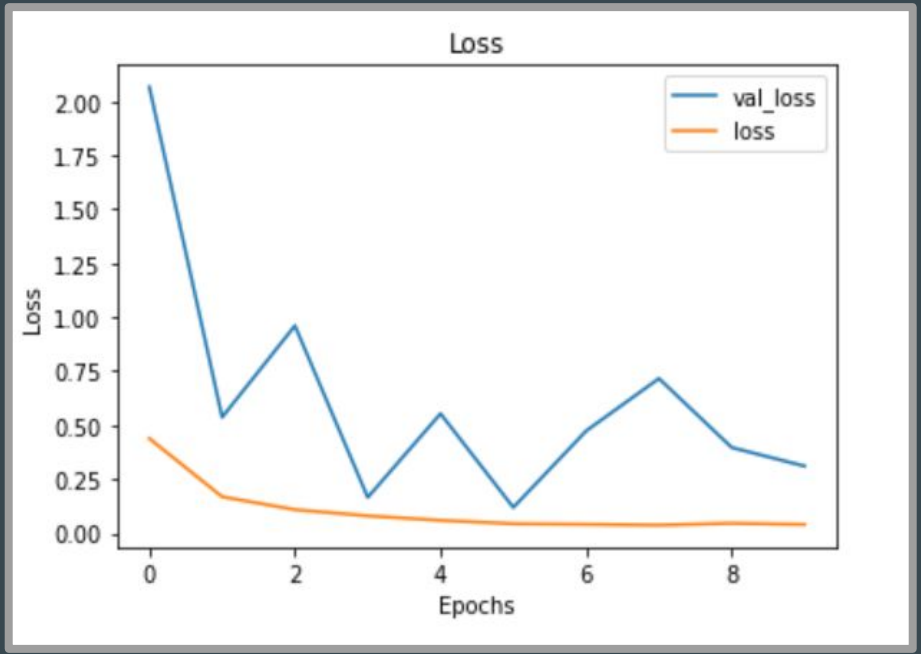




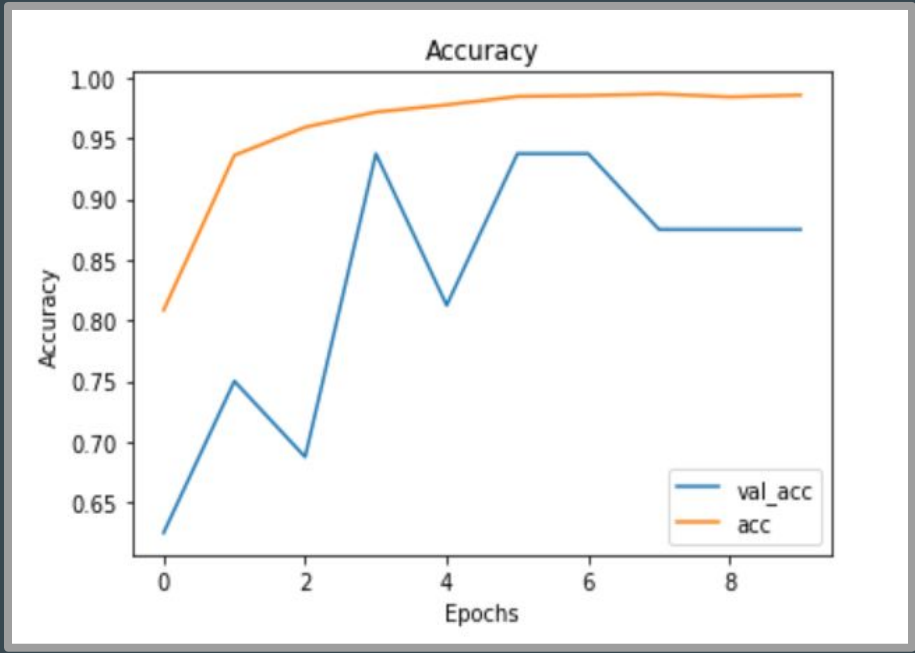
- Validation Loss: 0.29
- Training Loss: 0.16

- Validation Accuracy: 0.93
- Training Accuracy: 0.93

Model 3 (CNN + Added Layers)

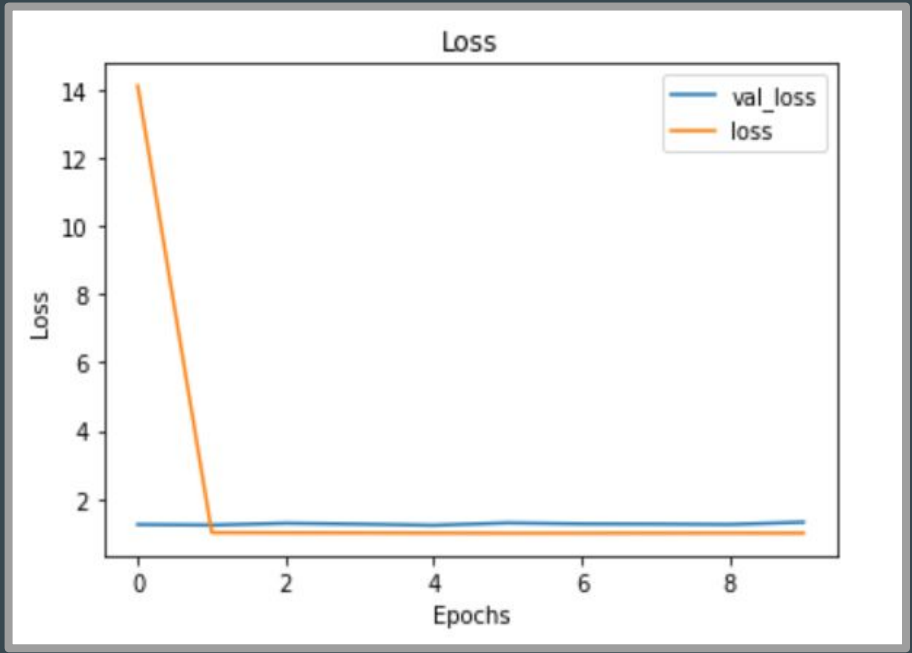


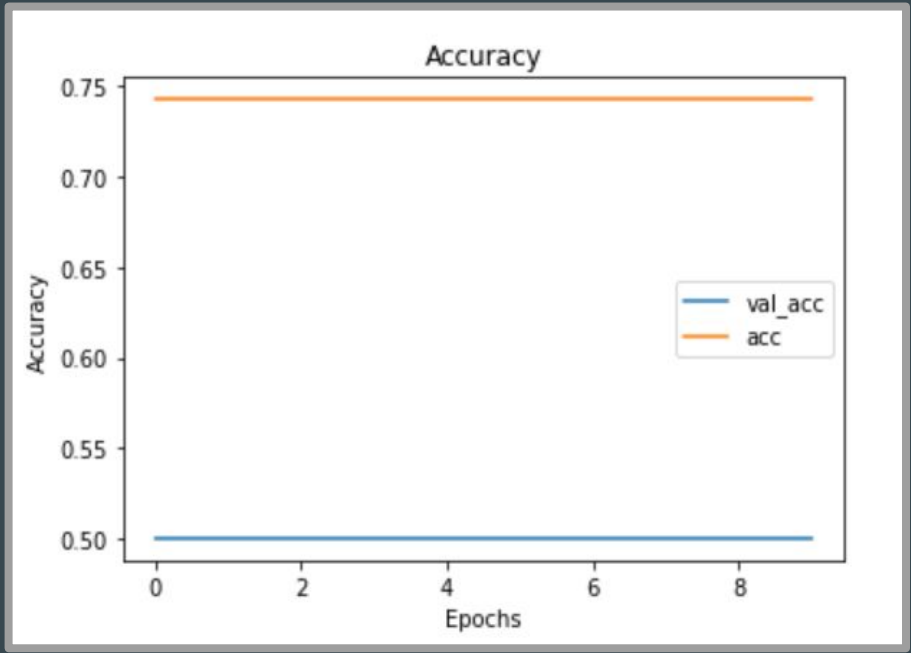
- Validation Loss: 0.31
- Training Loss: 0.01



- Validation Accuracy: 0.87
- Training Accuracy: 0.99

Model 4 (CNN + Layers + Regularization)





- Validation Loss: 1.00
- Training Loss: 1.32

- Validation Accuracy: 0.50
- Training Accuracy: 0.74

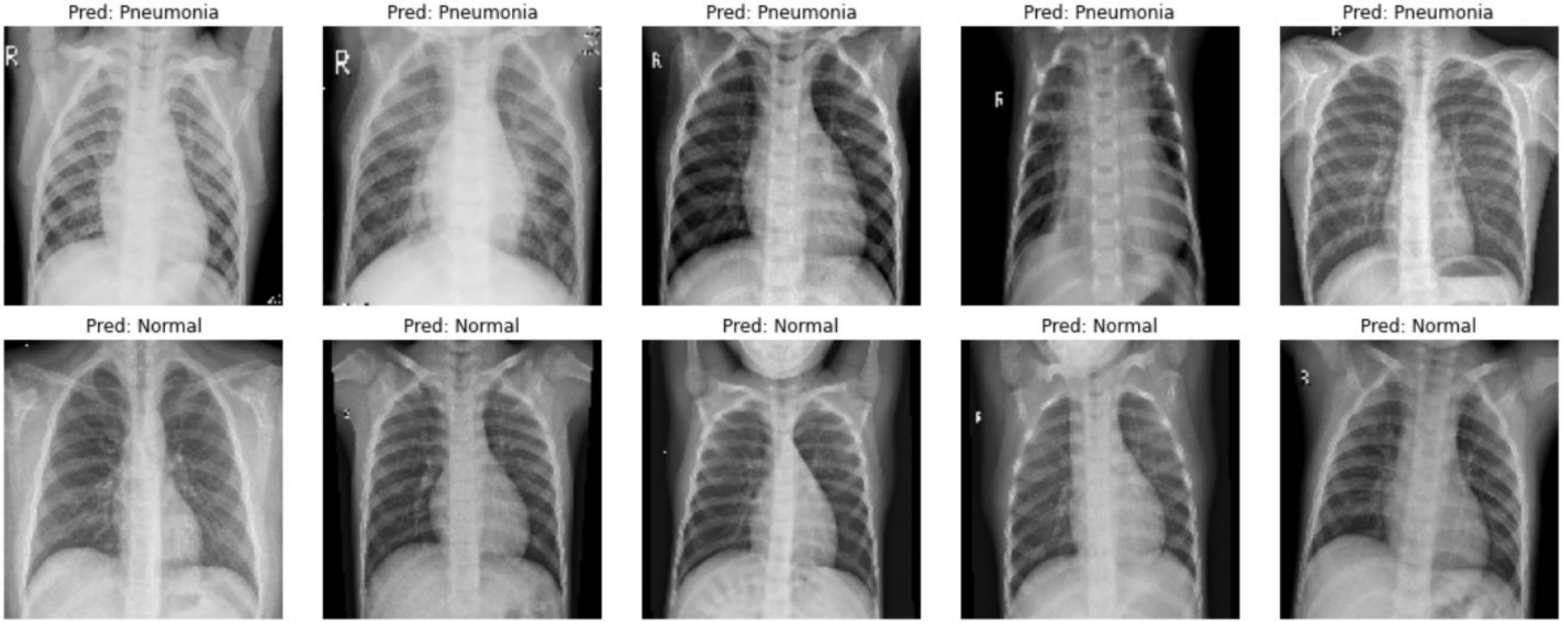
Evaluation

- Best Model = Model 2
- Other models had poor validation loss/accuracy score or were overfitting

TESTING LOSS: 0.41

TESTING ACCURACY: 0.81

Establish Predictions



Recall: 0.95

Precision: 0.79

F1-Score: 0.86

Conclusion

- Ultimately the best model for the Guangzhou Women and Children's Medical Center to use would be the Model 2, the convolutional neural network model
- The 0.95 recall score means that there is a 95% chance that this model will be able to correctly identify a patient testing positive more pneumonia (true positive)

Questions?