Lung Disease X-ray Classifier

By: Greg Fatouras

Business Problem

- How can we leverage automation to reduce diagnostic time?
- How can we ensure early detection and accurate diagnosis of lung diseases like COVID-19 and pneumonia?
- How can we lower the costs associated with diagnosing lung diseases while ensuring timely and effective patient care?

Dataset

- Thousands of labeled X-ray images
  - Normal: 10,192 → 2,000
  - Covid-19: 3,616 → 2,000
  - Pneumonia: 1,345 → 1,345

Covid-19

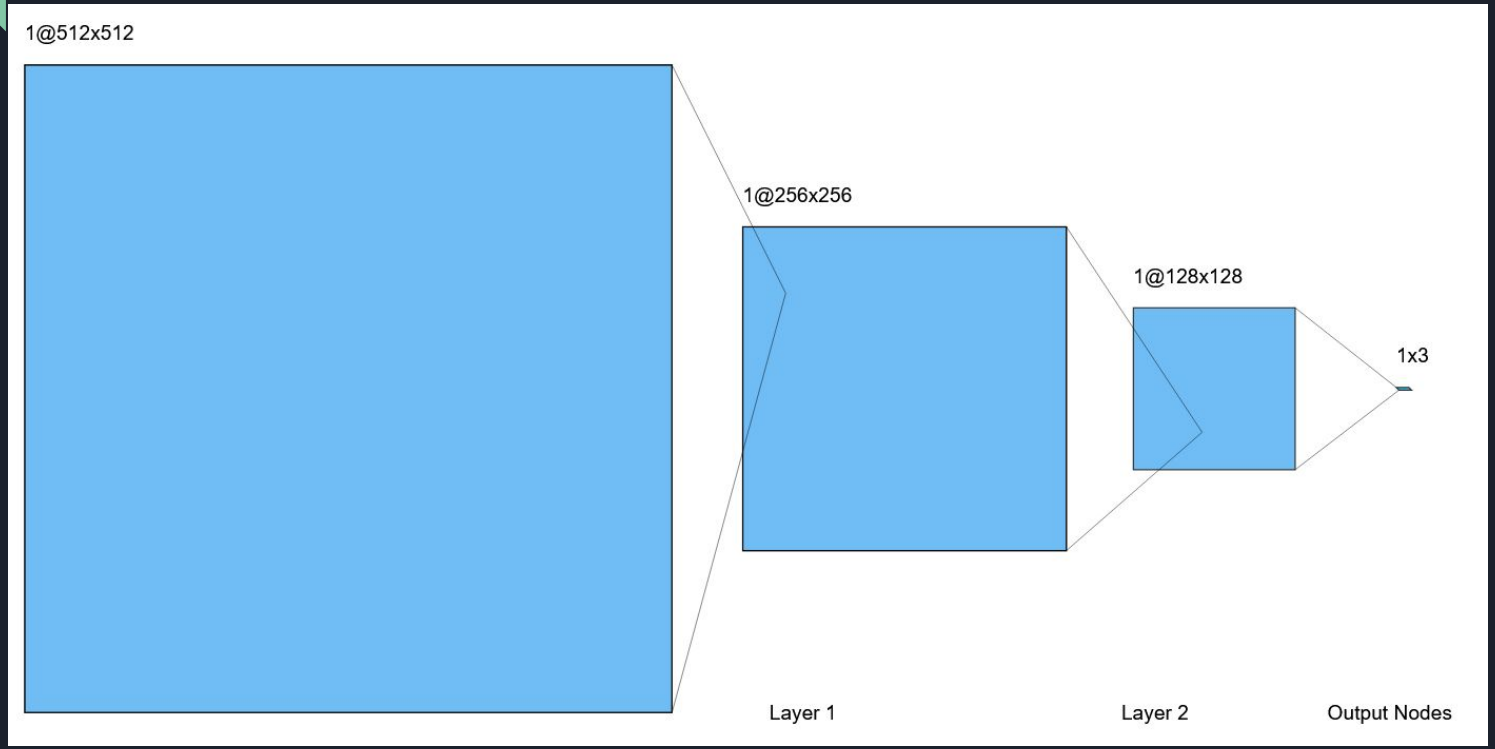
Pneumonia

Normal

Methods of tackling problem

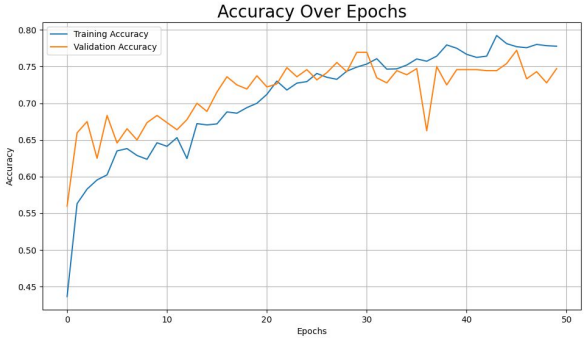
- Convert the 3 image classes to pixel arrays
- Apply Masks to X-ray images
- Create neural network to classify each image
- Hypertune parameters, include image manipulation

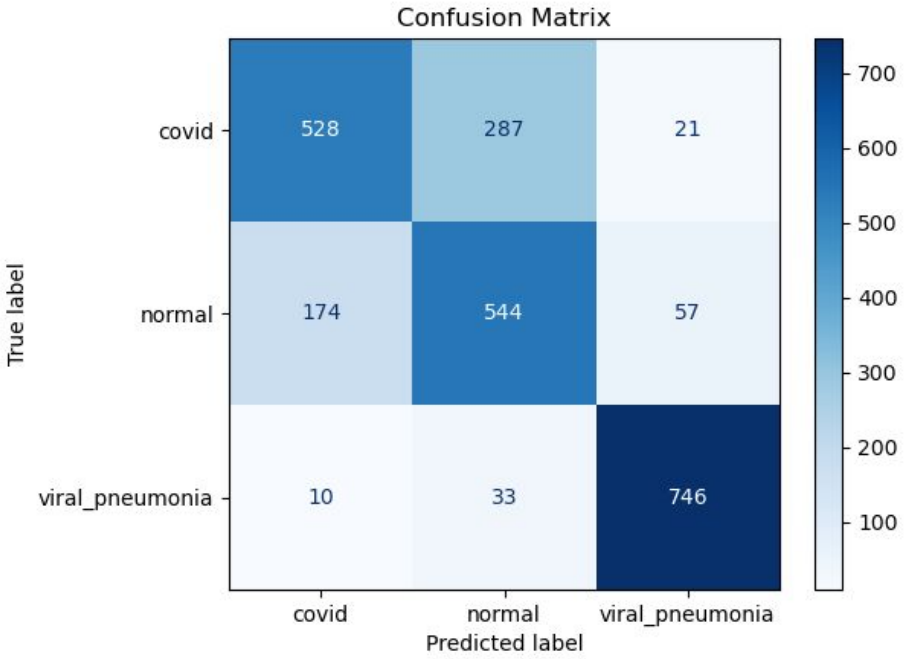
Baseline Architecture



Baseline Model

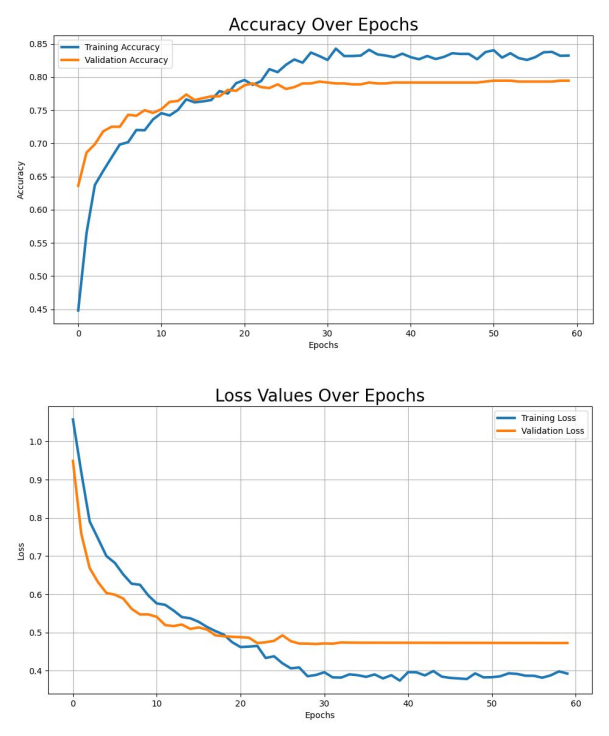
Train: 0.78 Test: 0.76

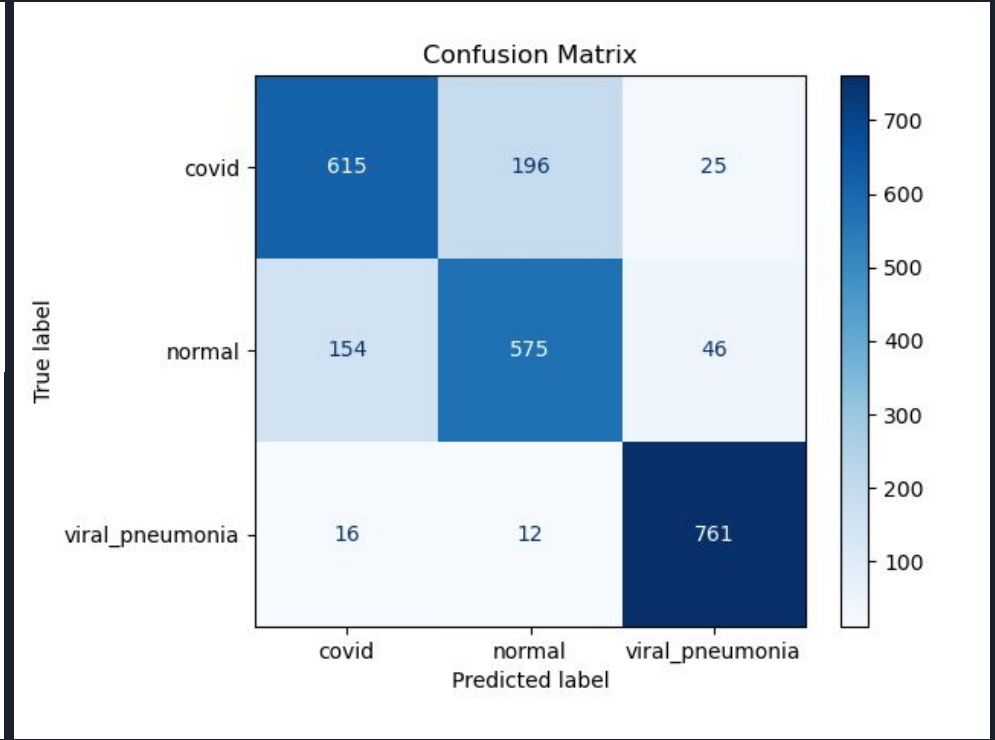




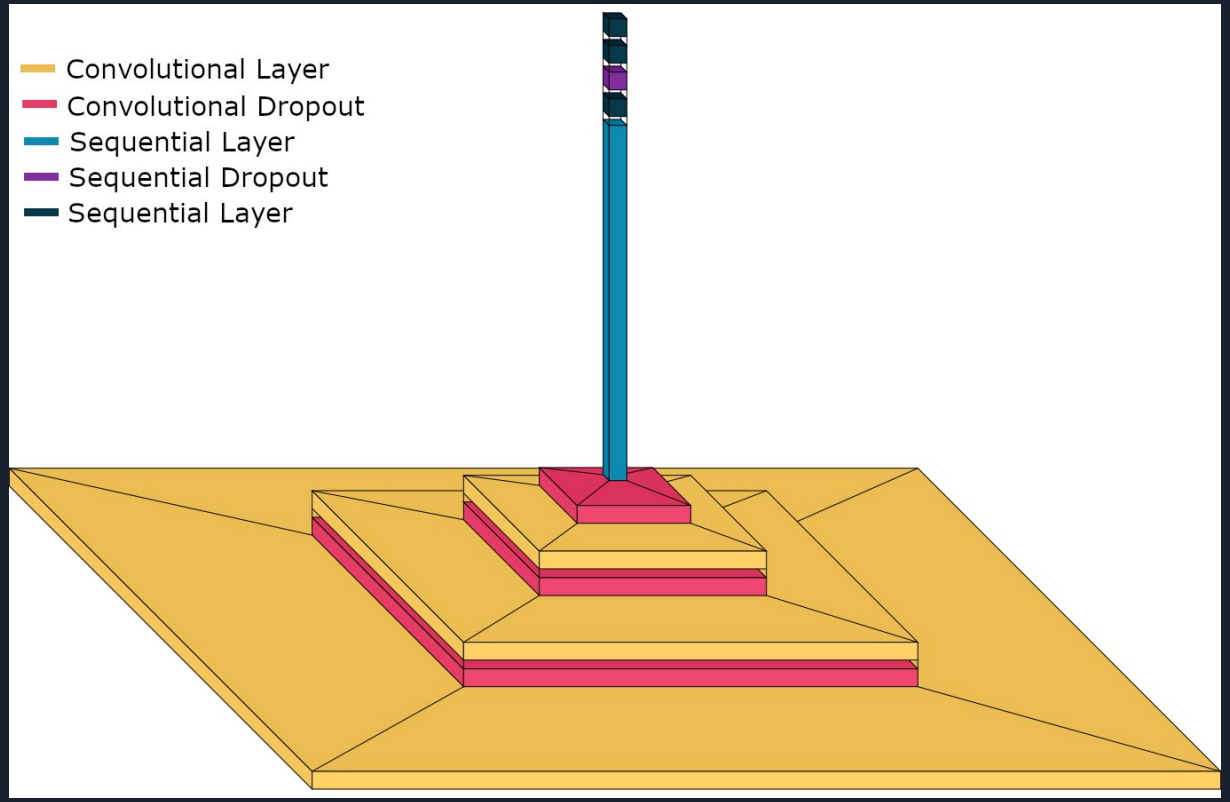
ReduceLROnPlateau

Train: 0.83 Test: 0.81



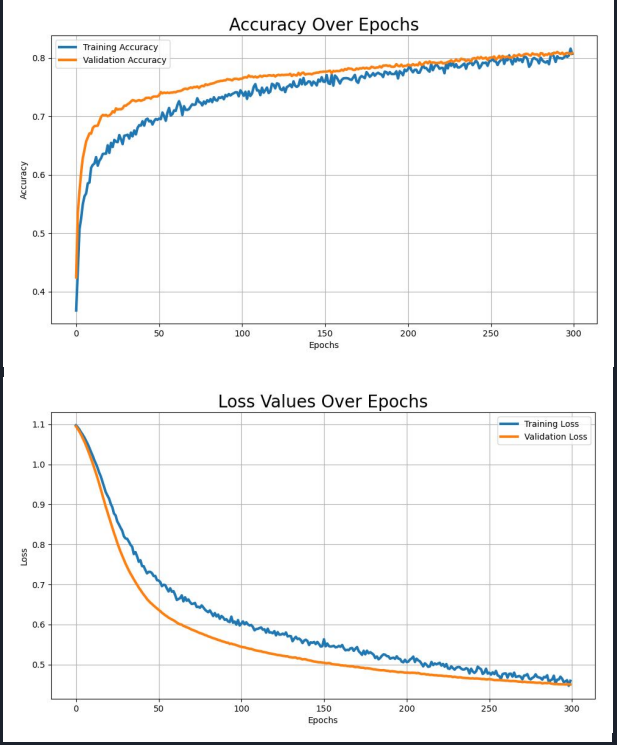


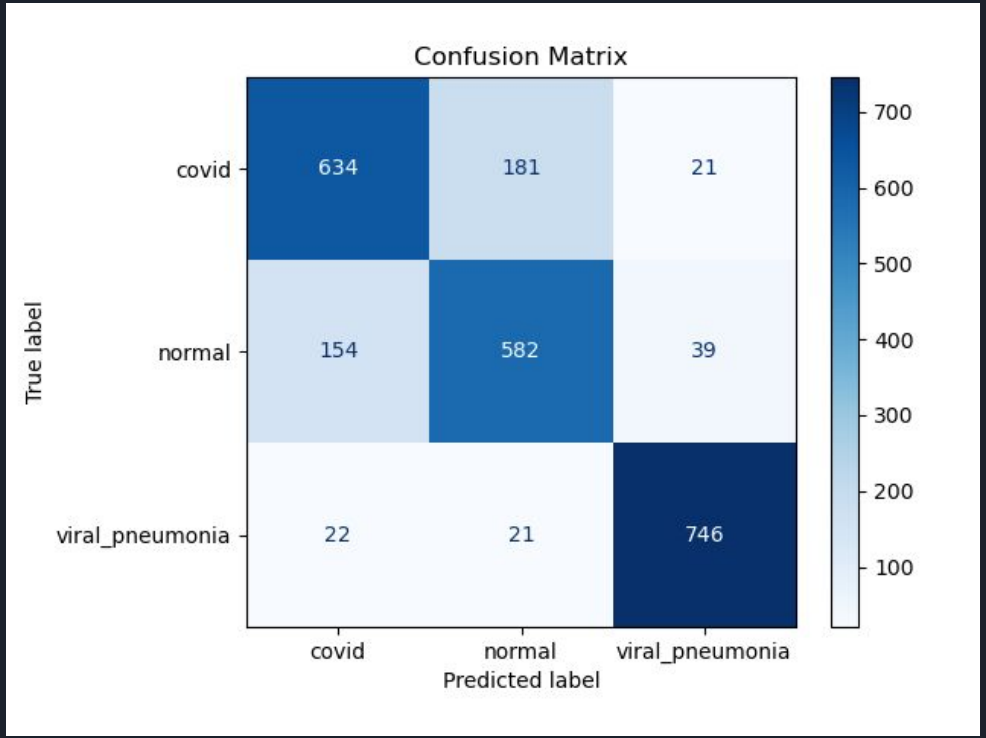
Convolutional Neural Network



CNN

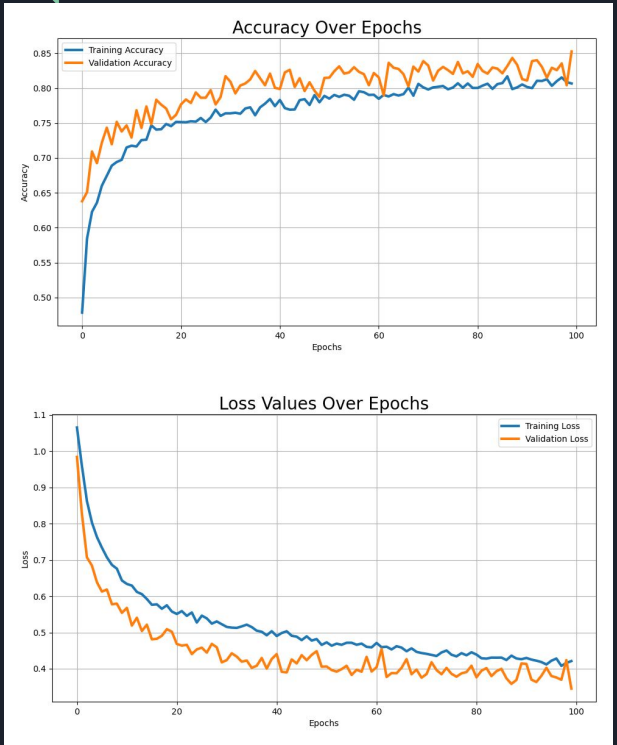
Train: 0.84 Test: 0.82

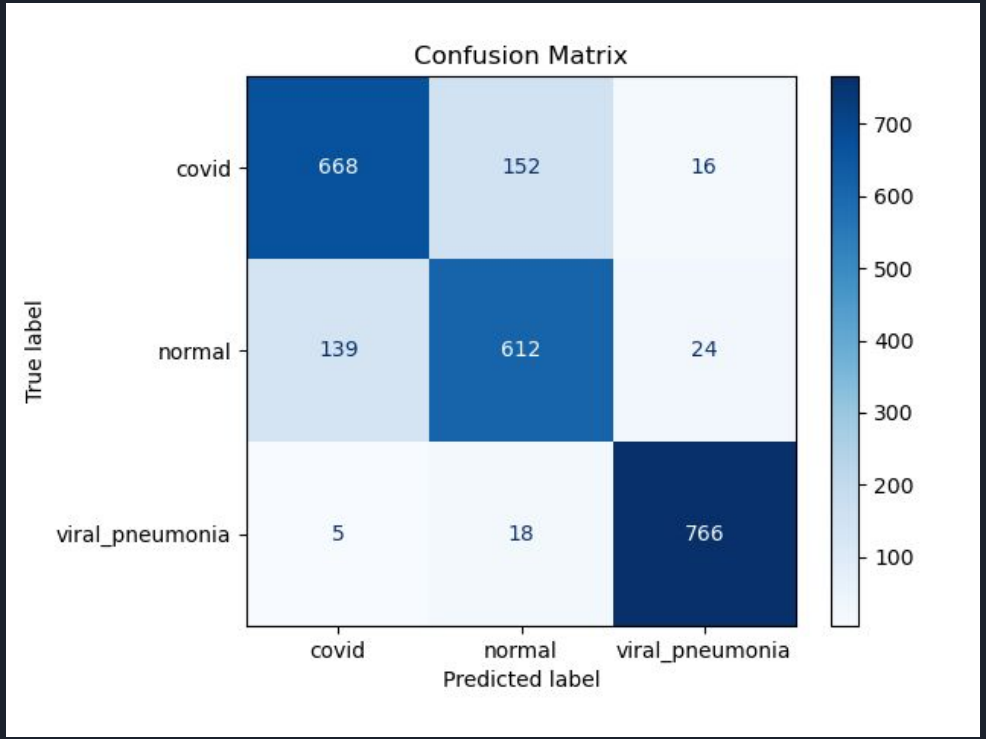




CNN + DataGen

Train: 0.81 Test: 0.86





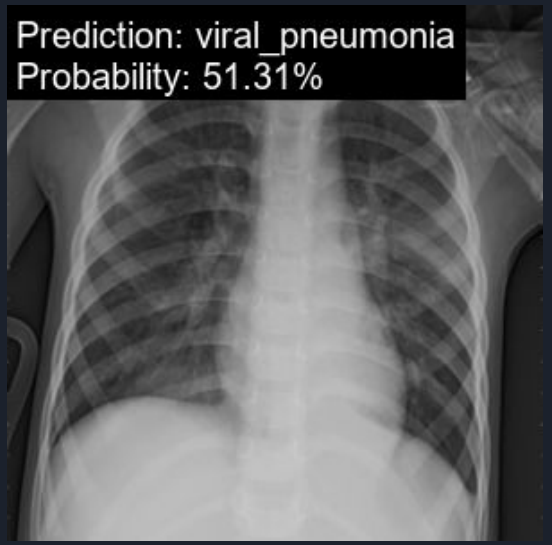
CNN + DataGen

Classification Report:

	precision	recall	f1-score	support
covid	0.85	0.78	0.81	836
normal	0.77	0.83	0.80	775
viral_pneumonia	0.96	0.97	0.97	789
accuracy			0.86	2400
macro avg	0.86	0.86	0.86	2400
weighted avg	0.86	0.86	0.86	2400

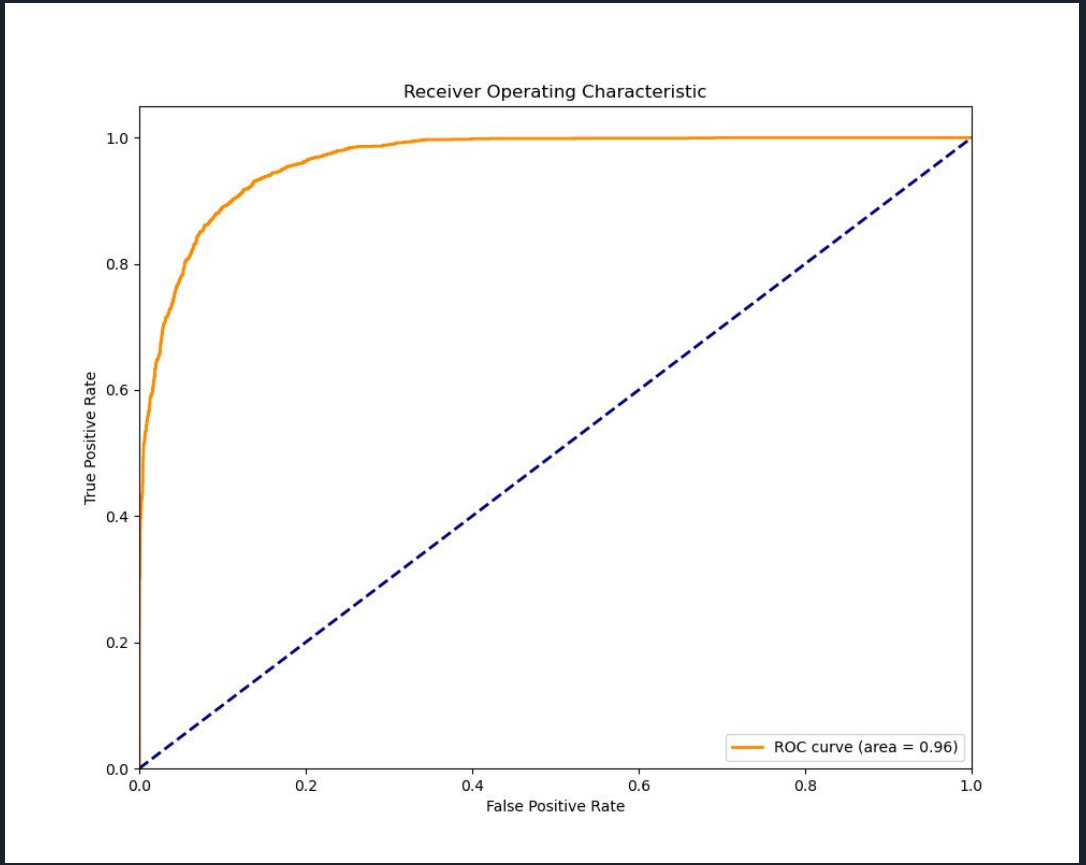
Image Classification

Pneumonia Example:



Conclusion

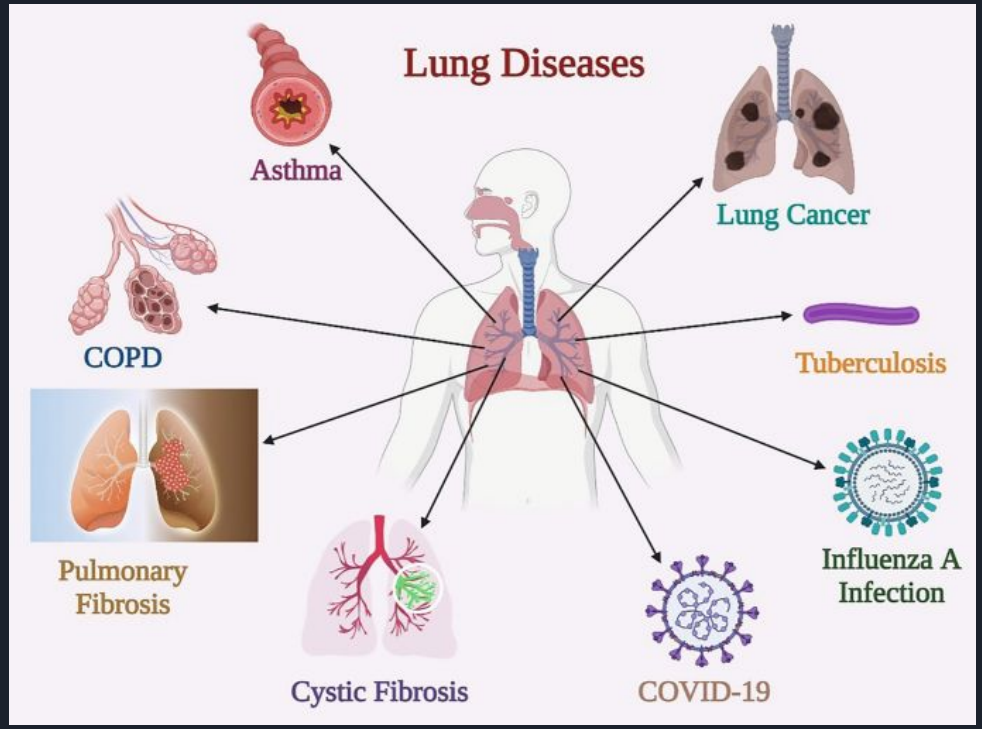
- CNN model with DataGen performed the best, with a test accuracy of 86%.
- Still struggles with normal/covid differentiation
- F1 scores of 80, 81, and 96 show a good balance between precision and recall.



CNN + DataGen ROC AUC curve

Future Steps

- Additional Normal and Covid data
- Using pre-trained CNN models
- Model evaluation and real-world testing
- Additional Lung disease images for greater classification coverage



Thank You

Greg Fatouras

- https://github.com/Gfatouras
- https://www.linkedin.com/in/gfatouras/
- fatourasg@gmail.com