



X-ray Pneumonia Classification

Avidan Berman & Dipta Roy

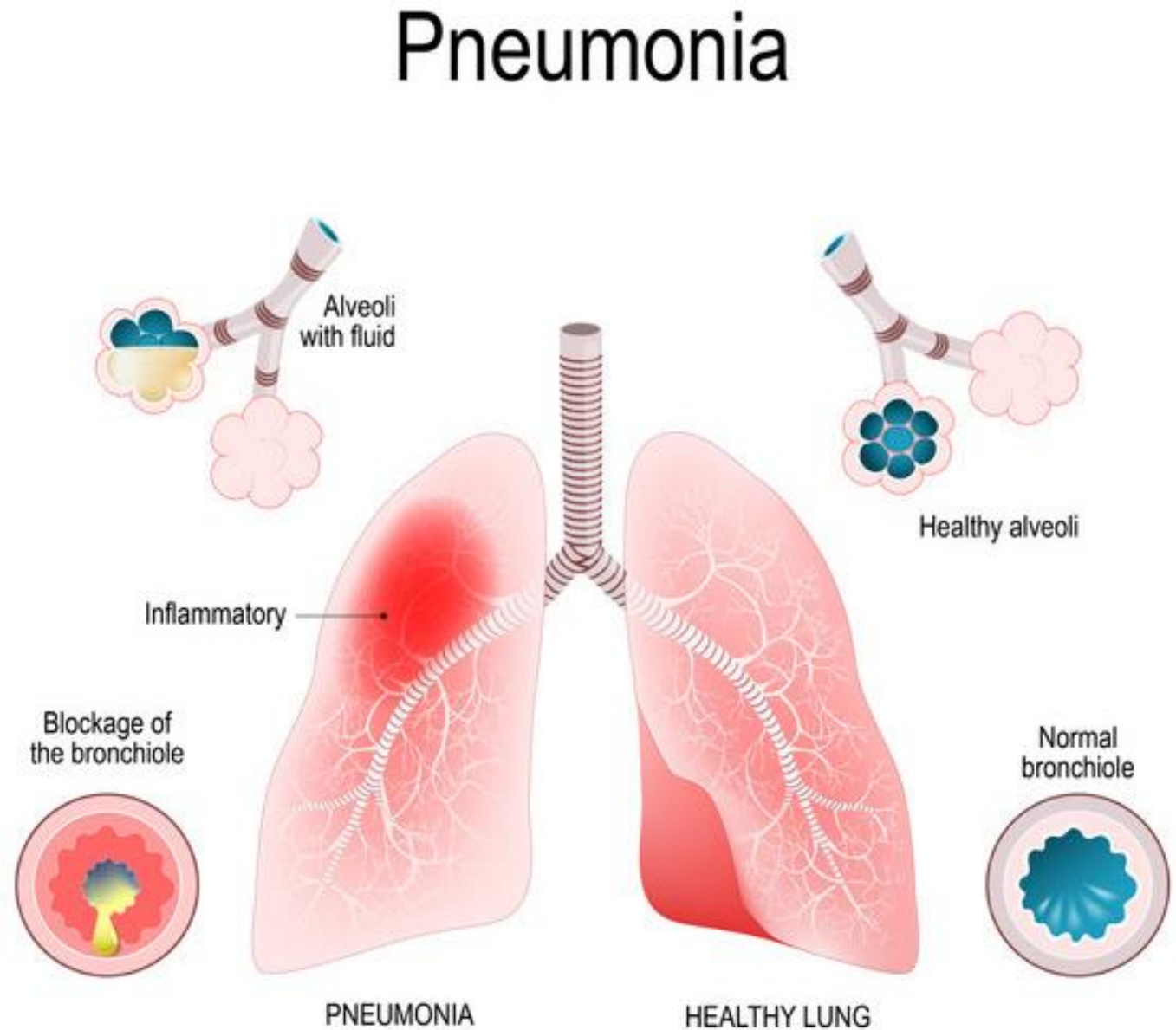


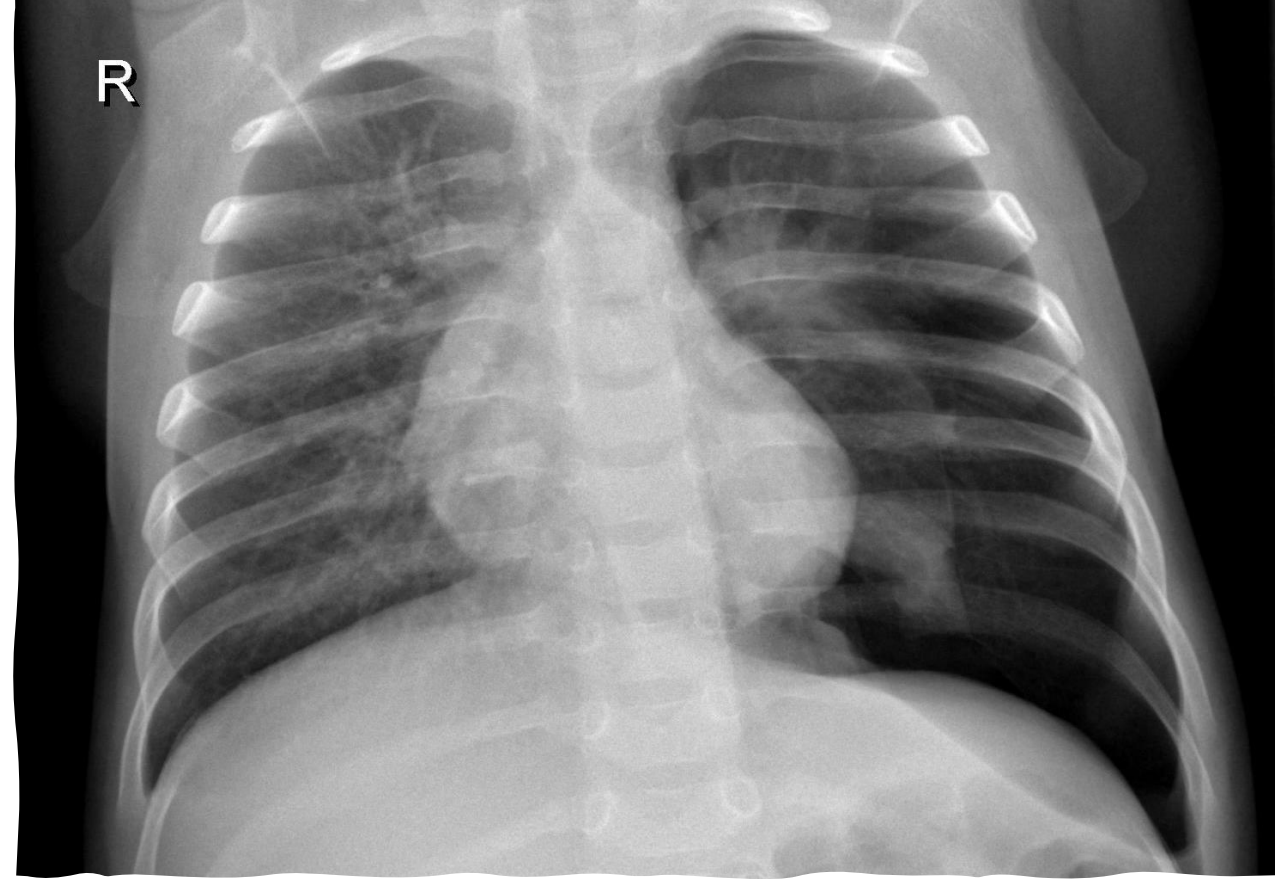
Overview

- Can we create an accurate model to predict if a person has pneumonia or not?
- Is our model scalable?
- Can our model be generalized?

What is Pneumonia

- **Pneumonia** is an infection that inflames the air sacs in one or both lungs. The air sacs may fill with fluid or pus, which causing cough with phlegm or pus, fever, chills, and difficulty breathing.
- A variety of organisms, including bacteria, viruses and fungi, can cause **pneumonia**



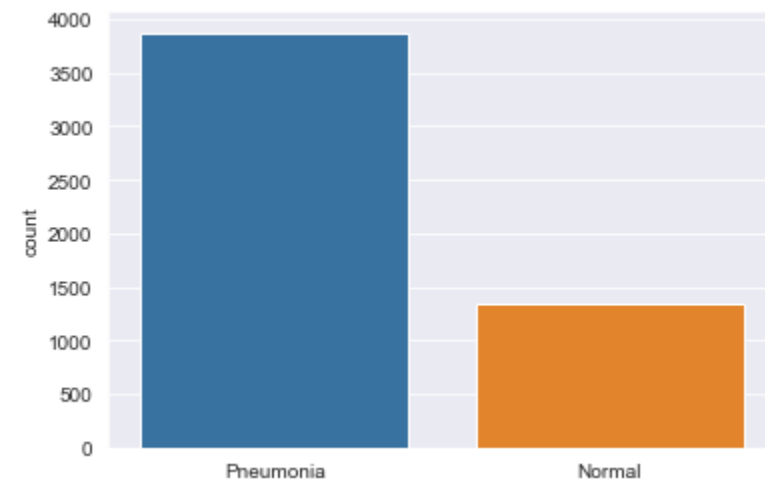


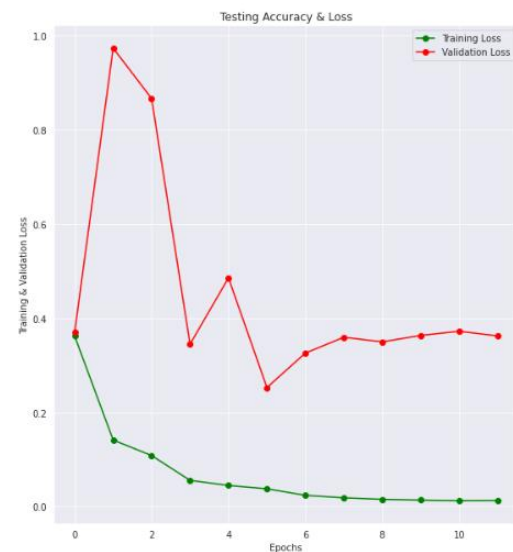
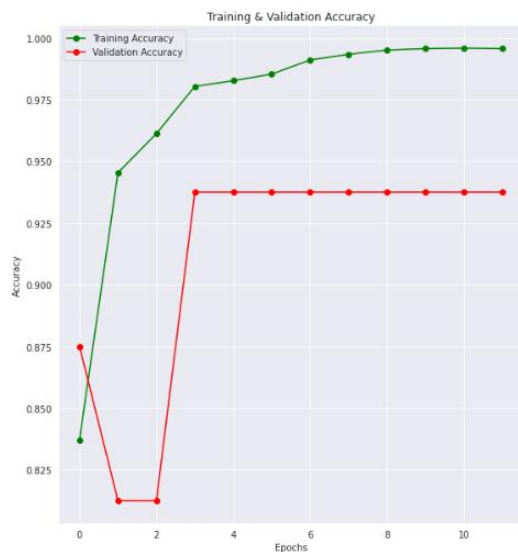
Can you tell the difference?

Which person has Pneumonia, the left or right?

Data Collection

- Dataset from:
 - Kaggle: [Chest X-Ray Images \(Pneumonia\)](#)
 - To offset imbalance: [CXR Data](#)
 - Testing sets:
 - [Chest X-Ray Images \(Pneumonia\)](#) -Test Set
 - [COVID-19 Xray Dataset](#)





	precision	recall	f1-score	support
Pneumonia (Class 0)	0.71	1.00	0.83	390
Normal (Class 1)	1.00	0.33	0.50	234
accuracy			0.75	624
macro avg	0.86	0.66	0.66	624
weighted avg	0.82	0.75	0.71	624

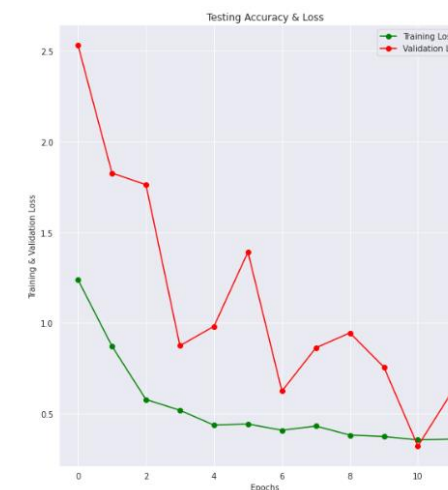
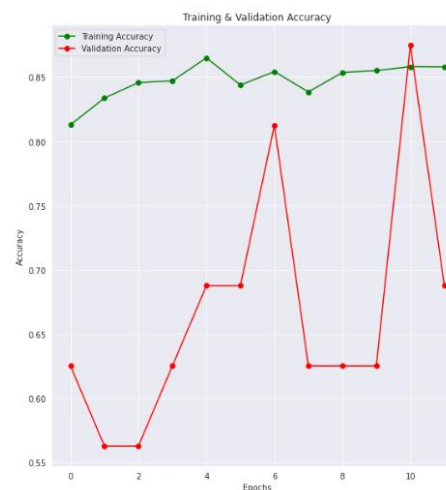
Baseline Model

- No fixing class imbalances
- CNN with 5 convolution layers and 1 dense layer

Dealing with class imbalance

- Data Augmentation

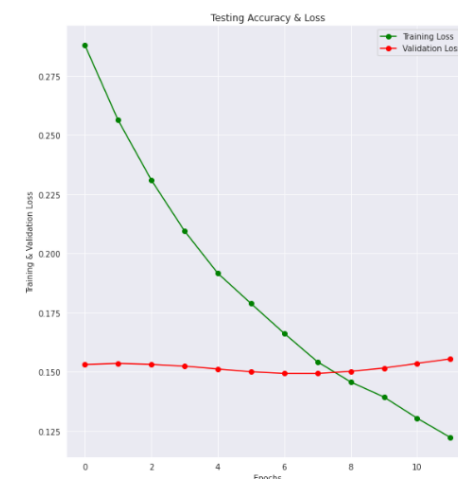
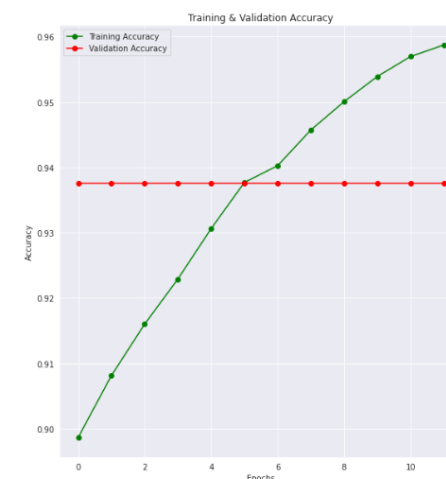
	precision	recall	f1-score	support
Pneumonia (Class 0)	0.81	0.97	0.88	390
Normal (Class 1)	0.93	0.61	0.74	234
accuracy			0.84	624
macro avg	0.87	0.79	0.81	624
weighted avg	0.85	0.84	0.83	624



Data Augmentation did better than just adding datapoints, but both did better than our baseline

- Inputting other datasets

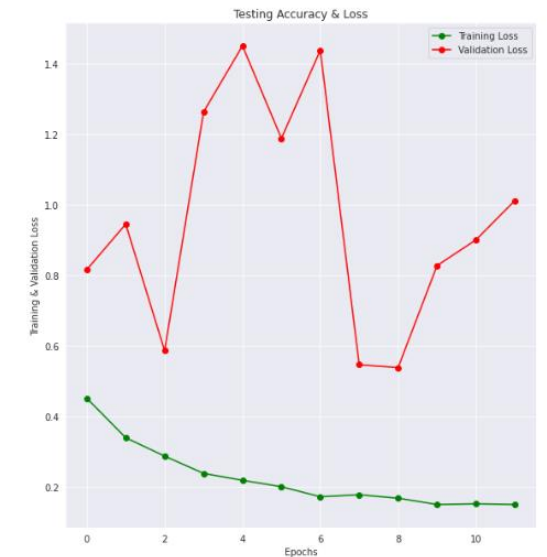
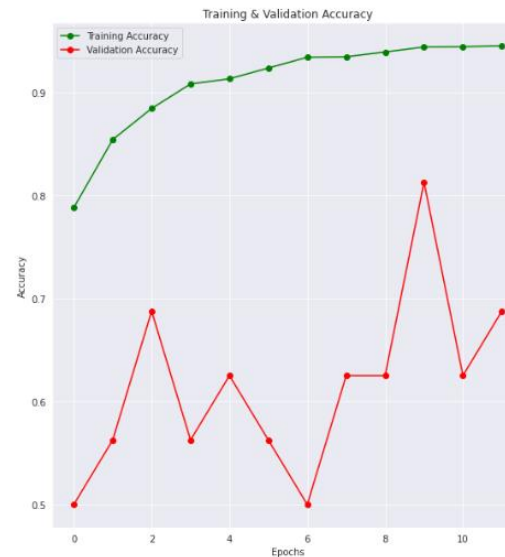
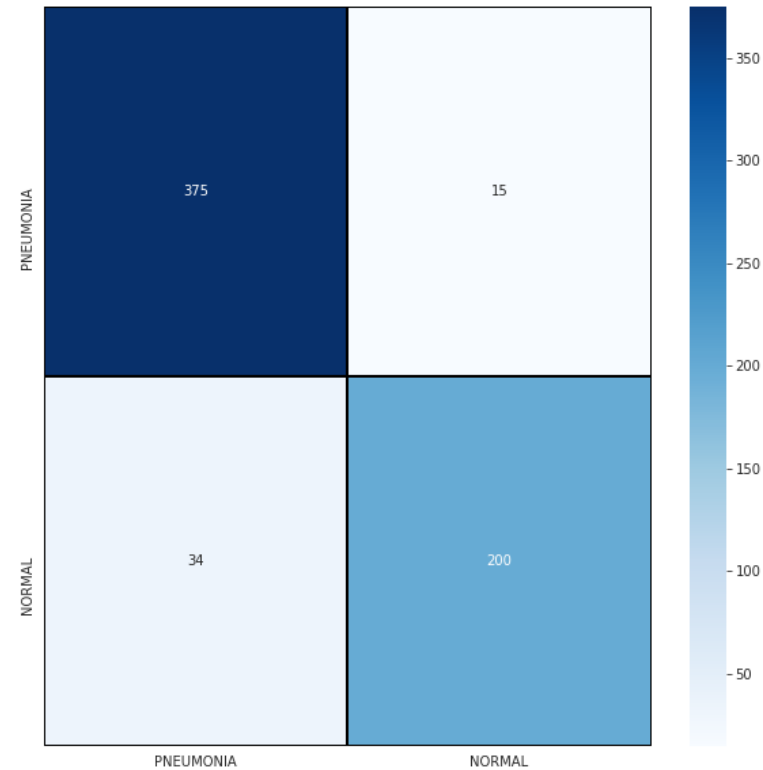
	precision	recall	f1-score	support
Pneumonia (Class 0)	0.77	0.99	0.87	390
Normal (Class 1)	0.98	0.52	0.68	234
accuracy			0.82	624
macro avg	0.88	0.76	0.77	624
weighted avg	0.85	0.82	0.80	624



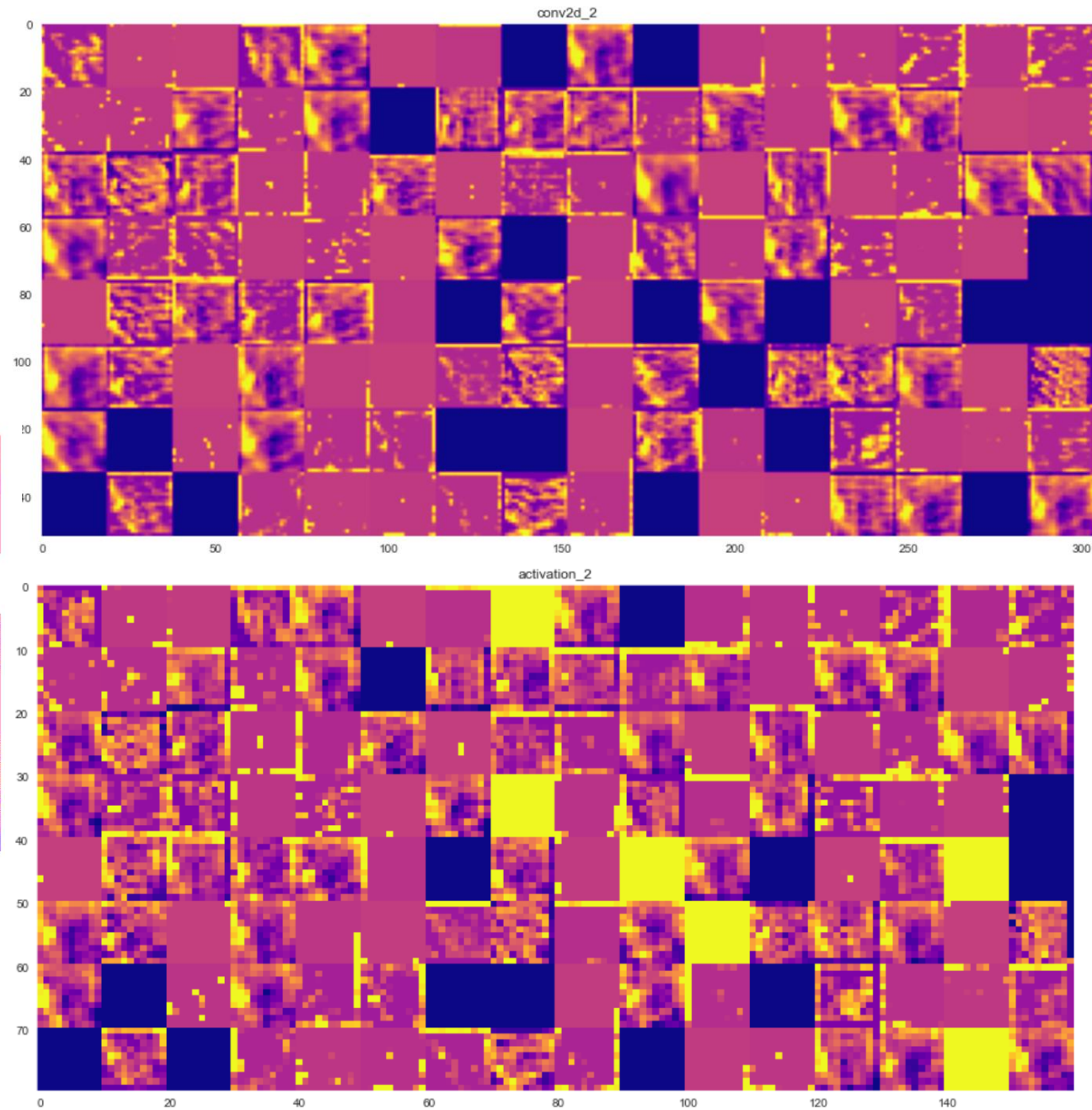
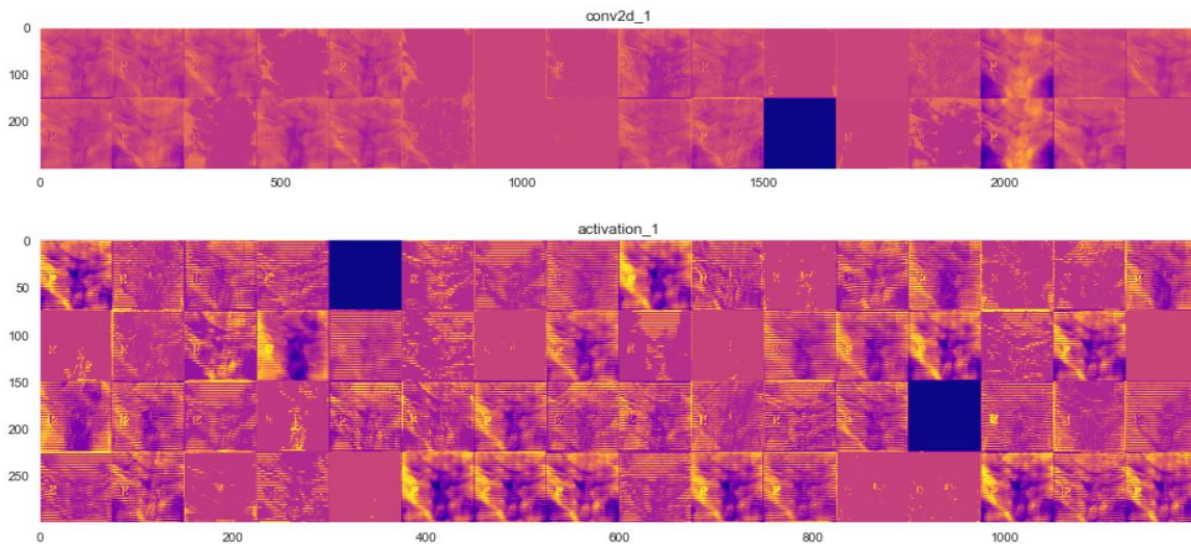
Final Model

- Includes extraneous data and data augmentation
- Increased Convolutions 4 to 6
- We cared more about FN's than FP's

	precision	recall	f1-score	support
Pneumonia (Class 0)	0.92	0.96	0.94	390
Normal (Class 1)	0.93	0.85	0.89	234
accuracy			0.92	624
macro avg	0.92	0.91	0.91	624
weighted avg	0.92	0.92	0.92	624



Let's visualize the layers



Future

- Based on our findings we have found that it is possible for CNN models to identify Pneumonia in patients.
- We recommend that doctors and hospitals move towards machine learning to identifying Pneumonia, this will save all party's time in the long run.
- In the future, we would like to expand our model to identifying other respiratory diseases.
- Lastly, we would like to create a front end to our model for medical professionals to use when attempting to identify Pneumonia.

Limitation

- Is bad at generalizing to non-clear x-rays, and over predicts to pneumonia
- Does not recognize other respiratory pathological issues, and may classify other symptoms as pneumonia.



*Thank you
and
Questions?*
