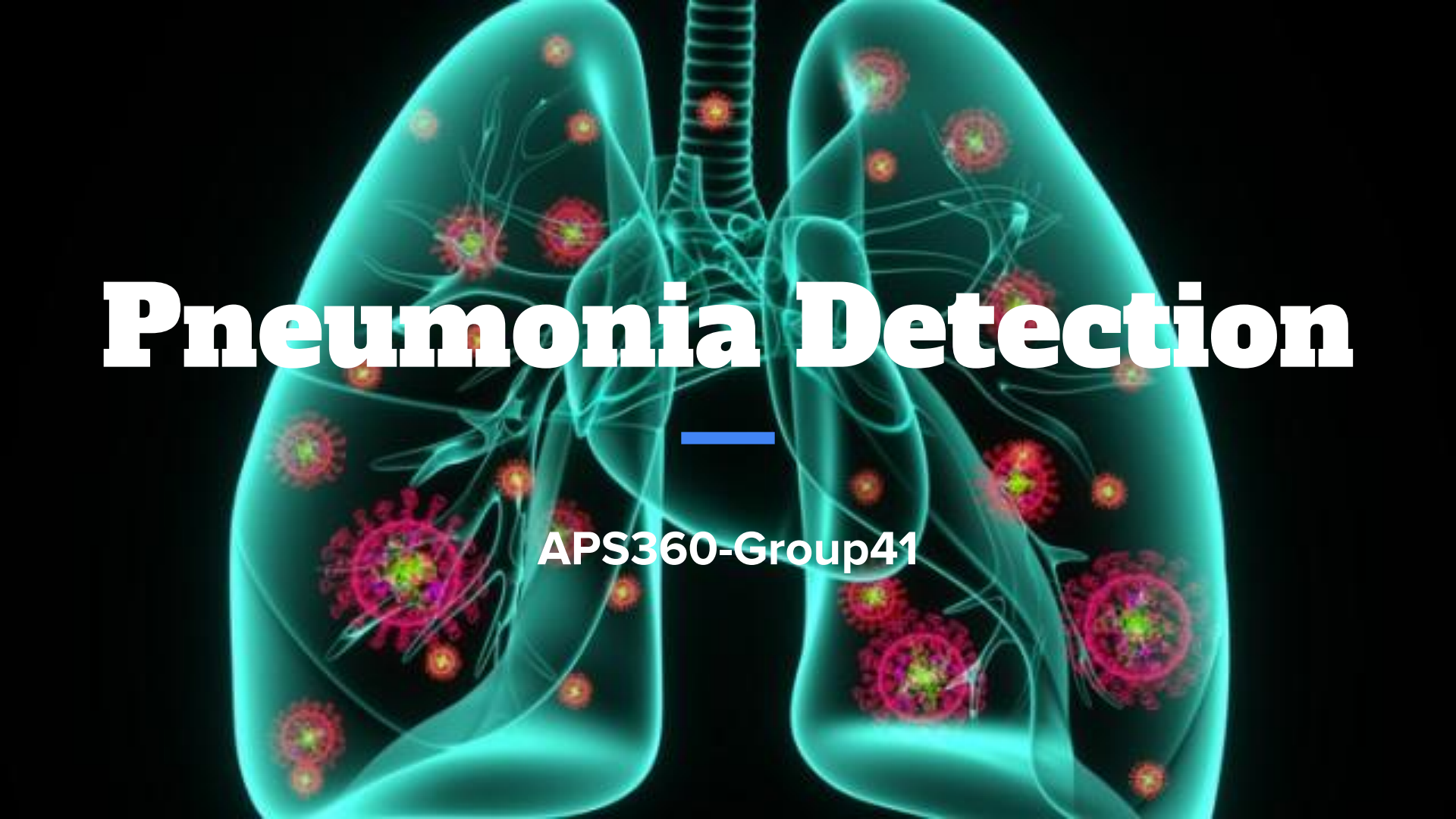
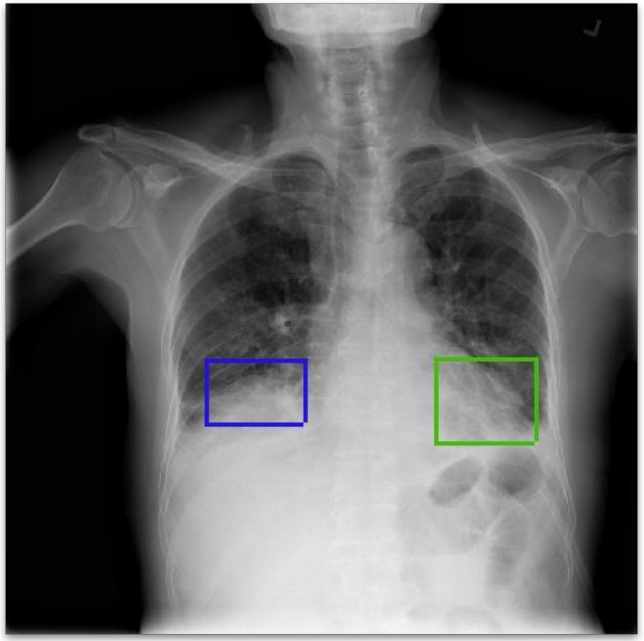


Pneumonia Detection

APS360-Group41

Pneumonia Detection

- Currently from chest scans, **physicians** are only able to correctly identify pneumonia with **47%** to **69%** accuracy
- Increasing Demand for **a more accurate and efficient diagnosis**

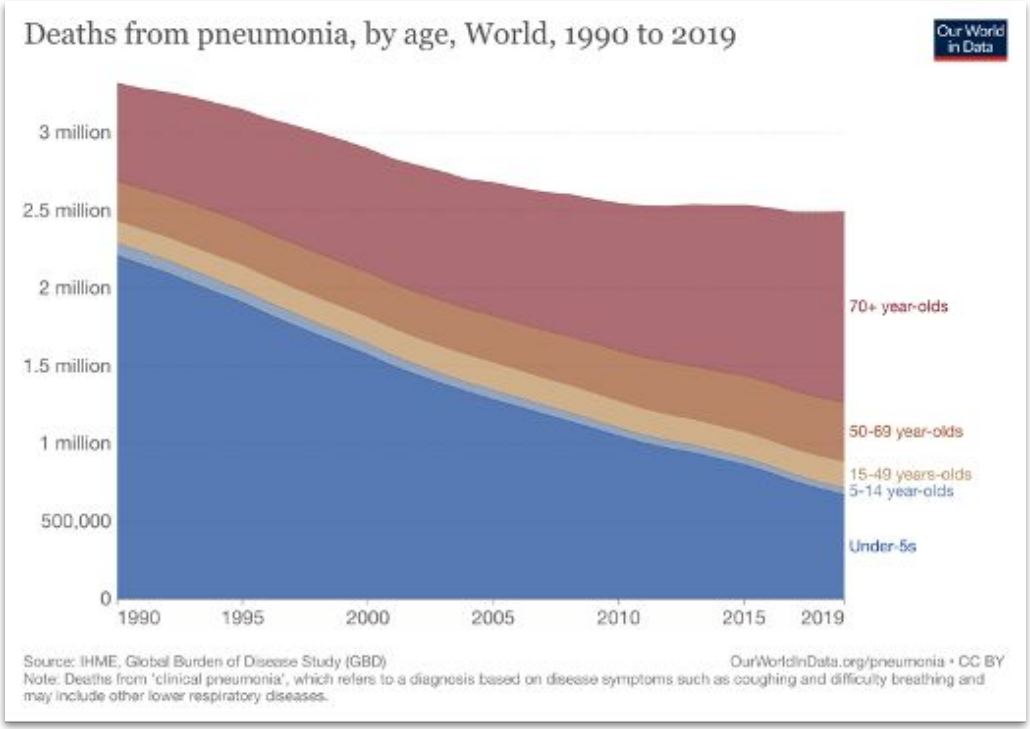


Our Model: 84.3%

Why is it important?

Goal: detect Pneumonia based on the Chest X-Ray

- most common diagnosis for severe **COVID-19** is severe pneumonia
- **24.8** cases per **1000** adults are diagonalized with pneumonia diseases in the United States
- **one of the top ten leading causes of death** in Canada



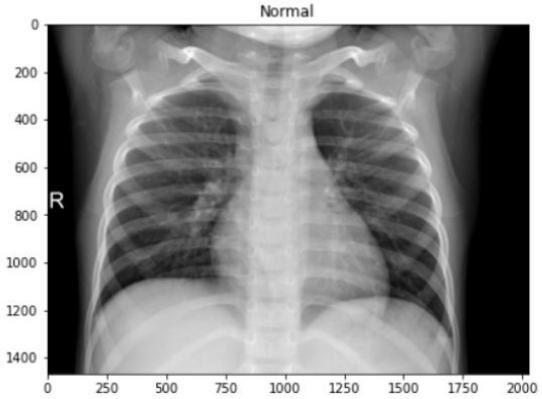
Data Collecting

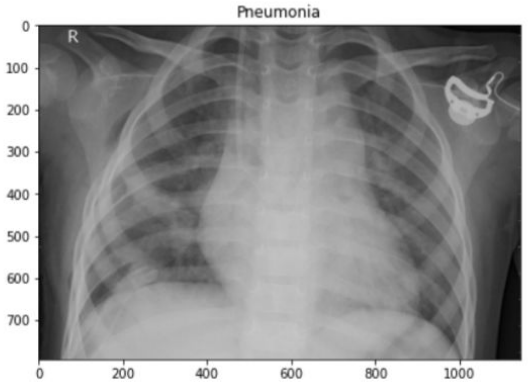
1

- Kaggle - "Chest X-Ray Images”
- 5,863 images
- Two classification labels
- Guangzhou Women and Children's Medical Center's retrospective cohorts of children aged one to five years old

2

Manual Quality Control





Loading Data into Notebook

3

4

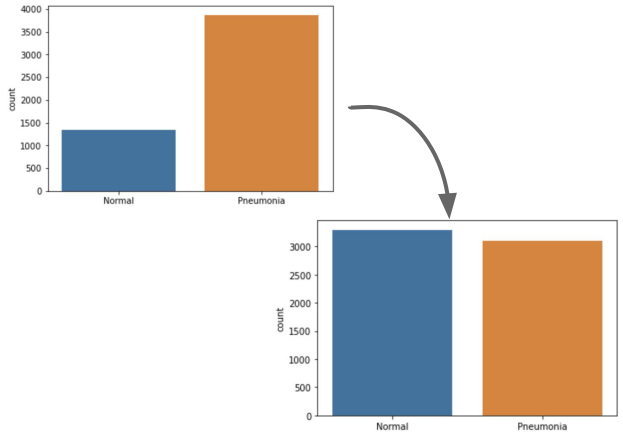
Train/Validation/Test Split

5216 images in training
16 in validation
624 in testing

→ 6:1:1

Training label Rebalance

5



6

Data Augmentation

- Horizontally flipping
- Shifting
- Shearing
- Zooming
- normalized

Choosing a Model

1. Sequence in Data → RNN

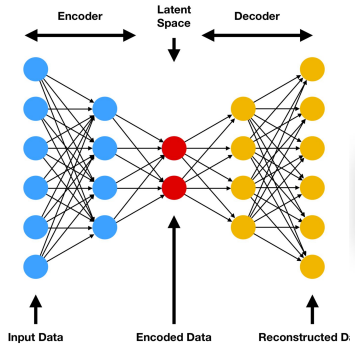
2. Image Recognition → CNN

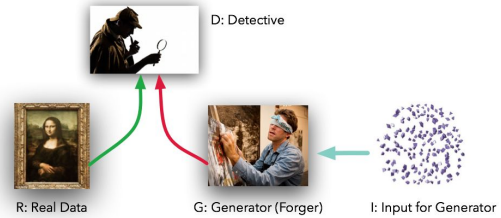
   Simple Classification

3. In the Future

Autoencoders

GANS





Model: CNN

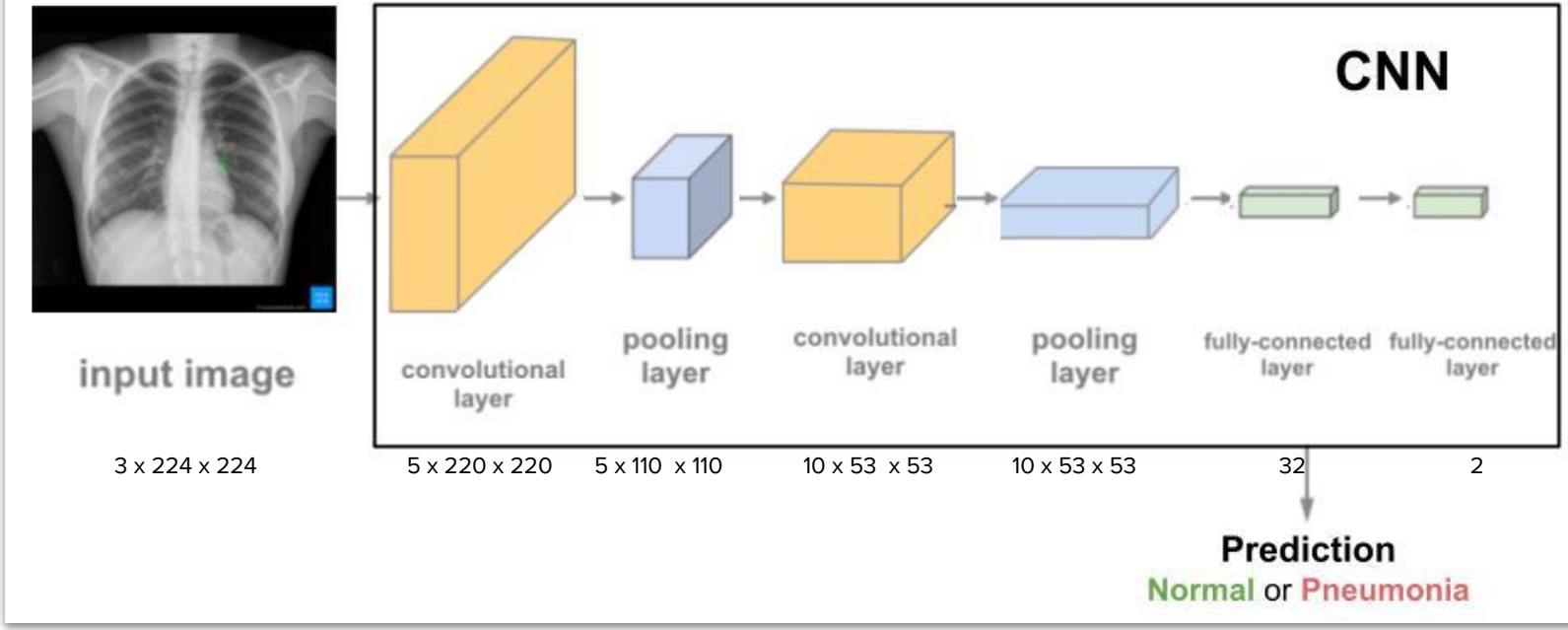
out_channels = 5
Kernel = 5 x 5
s = 1

Kernel = 2 x 2
s = 2

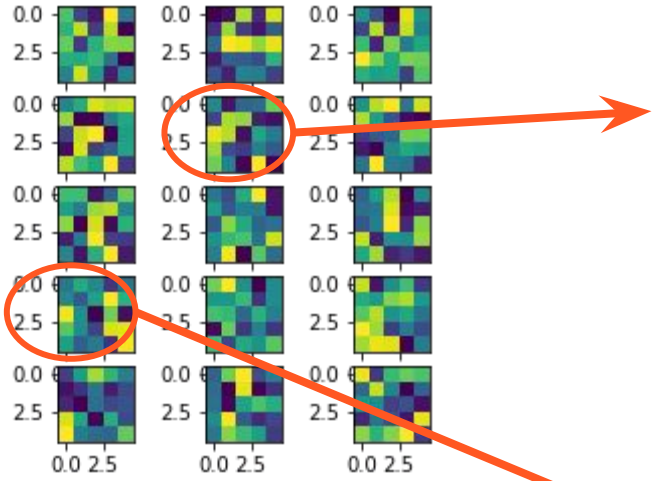
out_channels = 10
Kernel = 5 x 5
s = 1

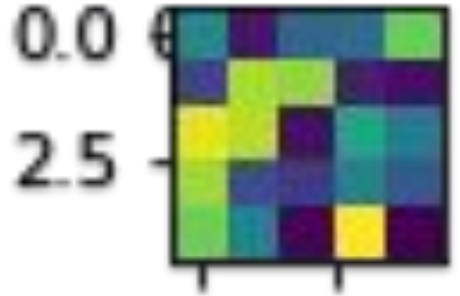
Kernel = 2 x 2
s = 2

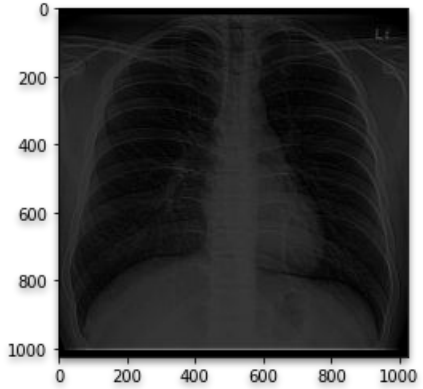
Flatten to
10 x 53 x 53

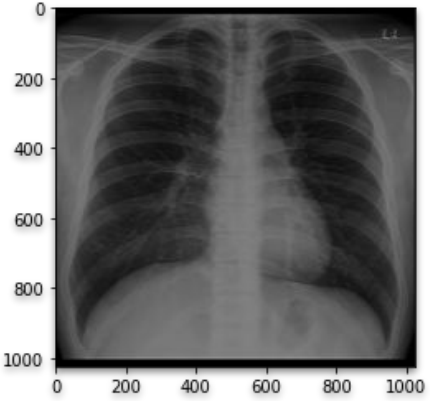


Kernel Visualization



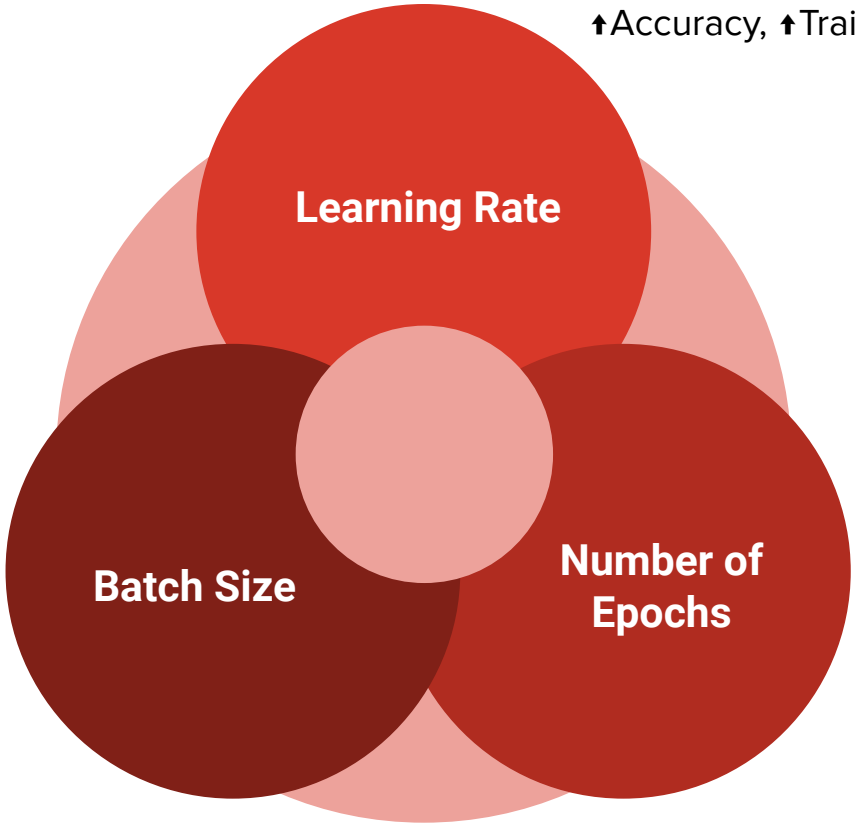






Hyperparameters

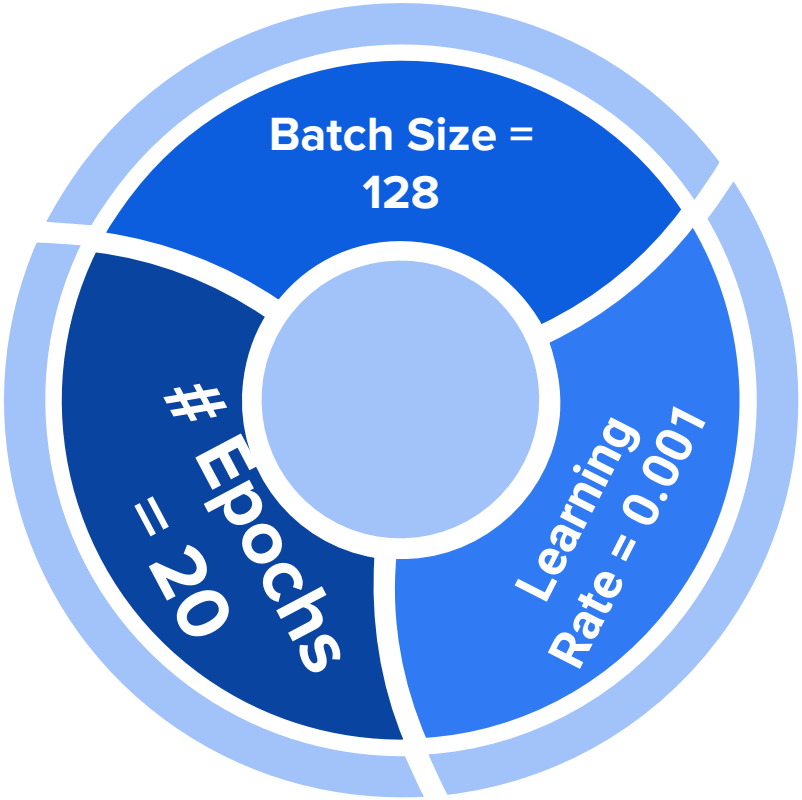
↓Learning Rate
↑Accuracy, ↑Training time



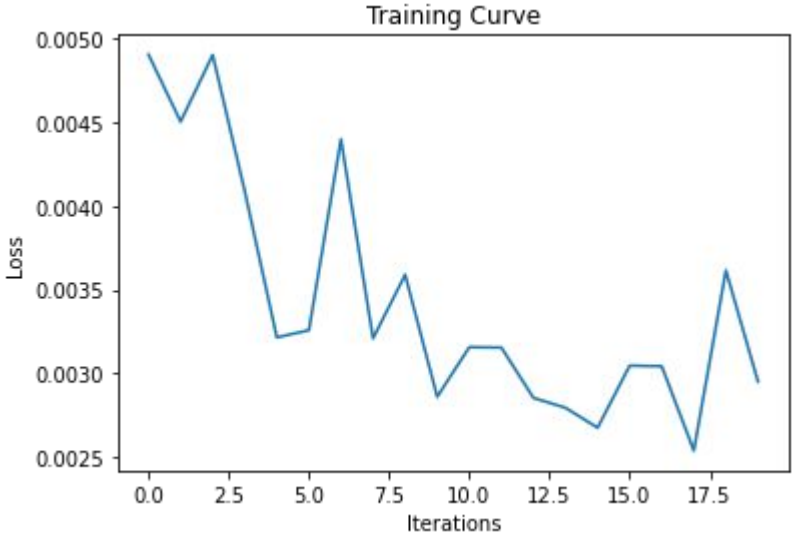
↓Batch Size
↑rate of convergence to “good” solutions

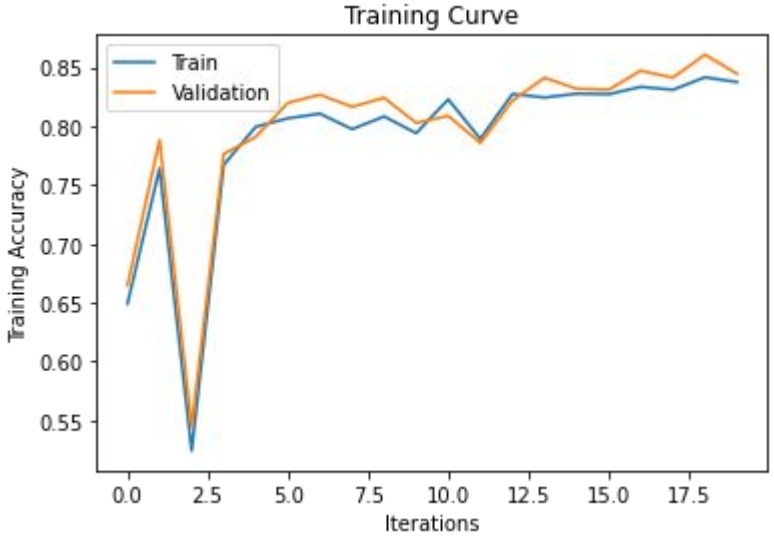
↑# of Epochs
↑Accuracy, ↑Training time
⚠️avoid overfitting

Final Classification Model

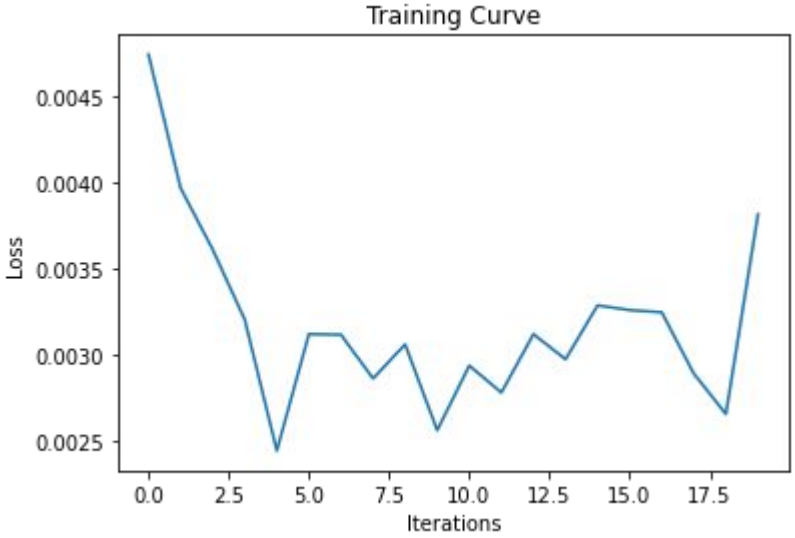


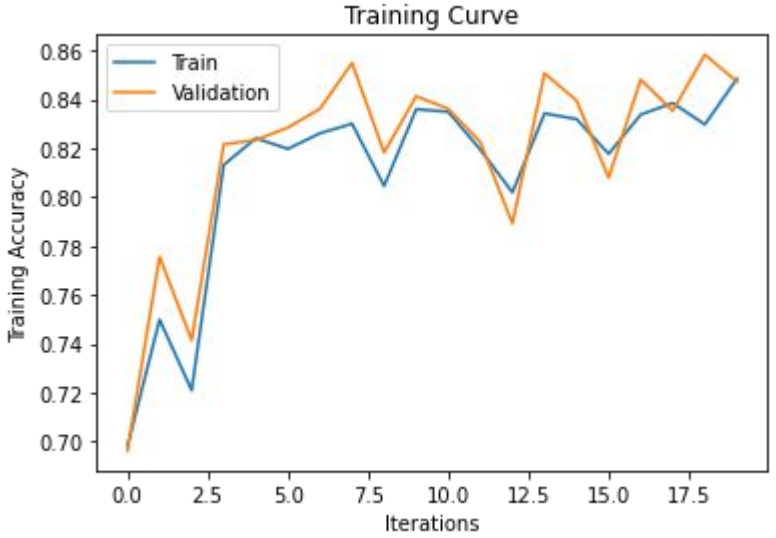
Validation Results



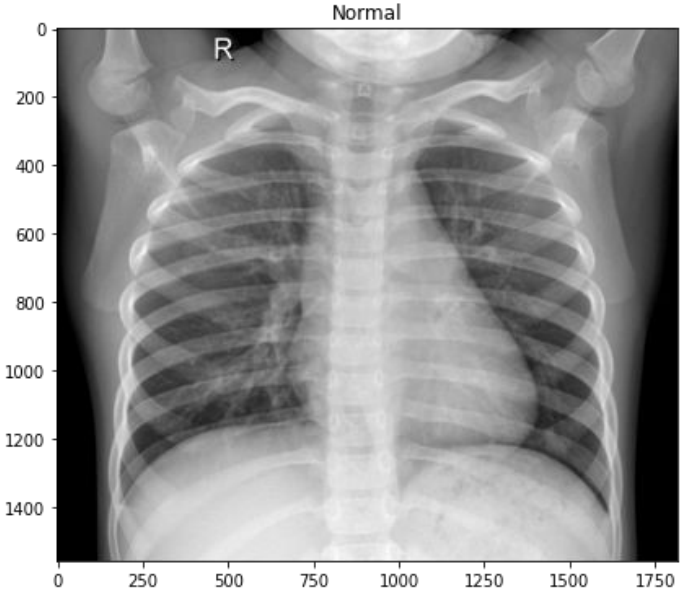


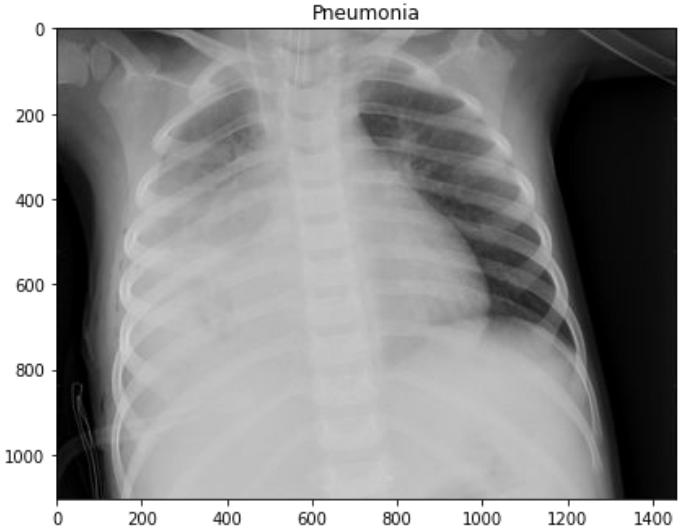
Testing Results





Sample Predictions





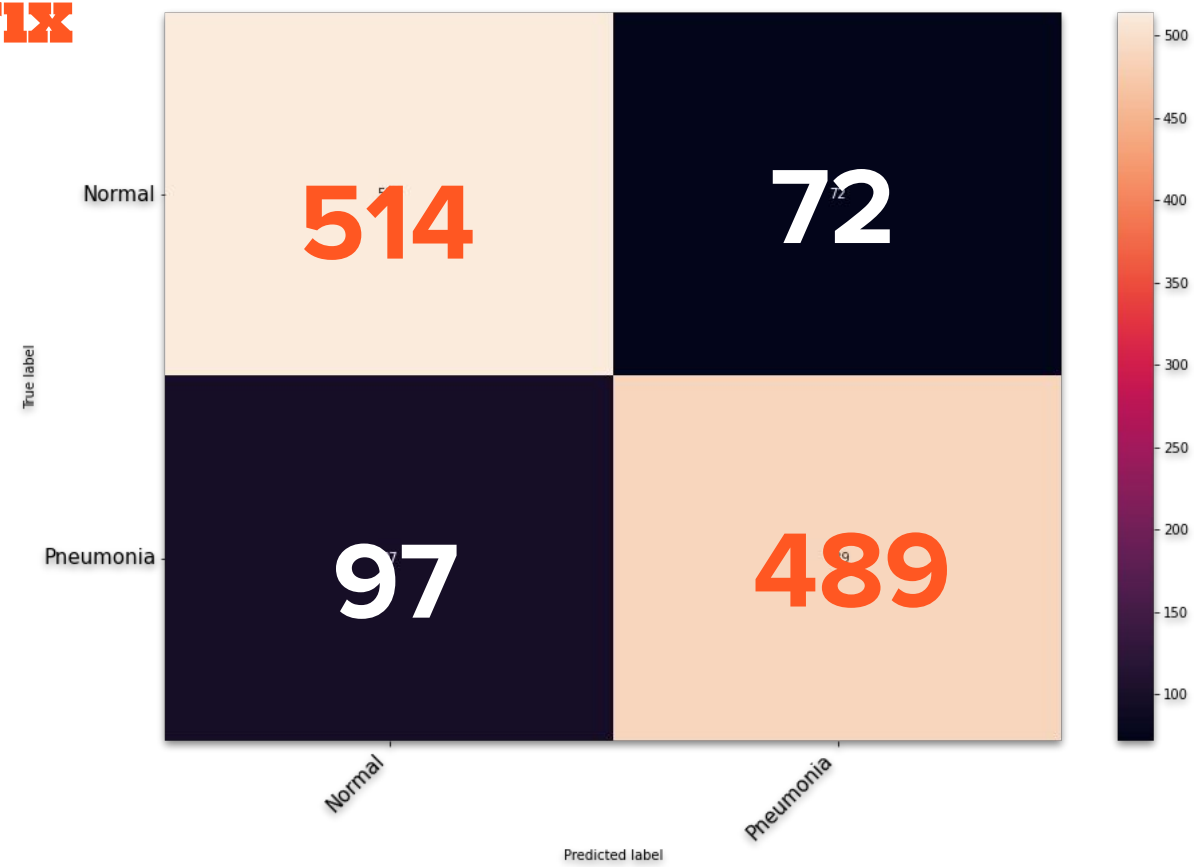
Model Prediction

87.5% **76.8%**

Confusion Matrix

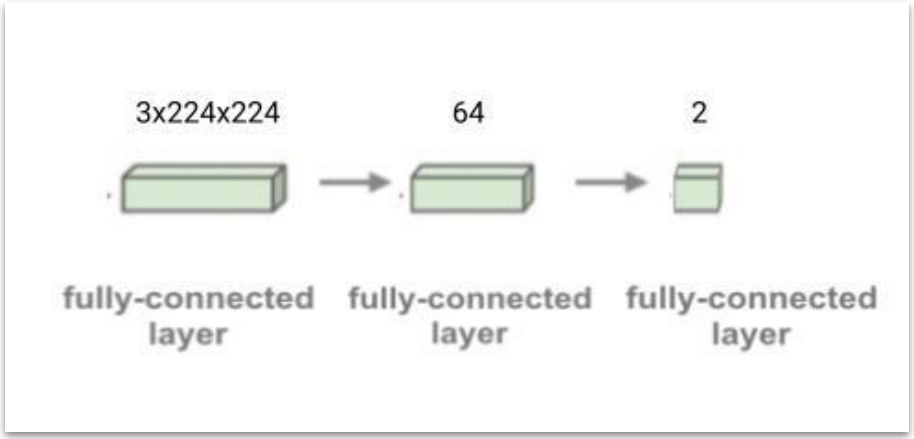
Recall should be maximized

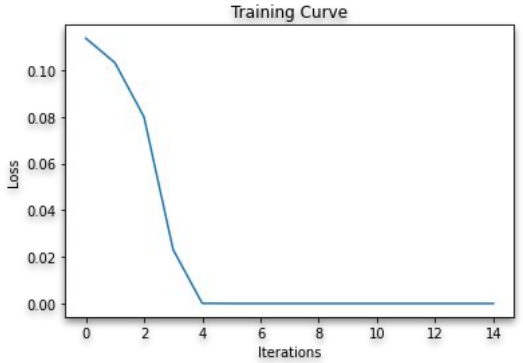
- **Precision** = 489/(489+72) = 0.87
- **Recall** = 489/(489+97) = 0.83

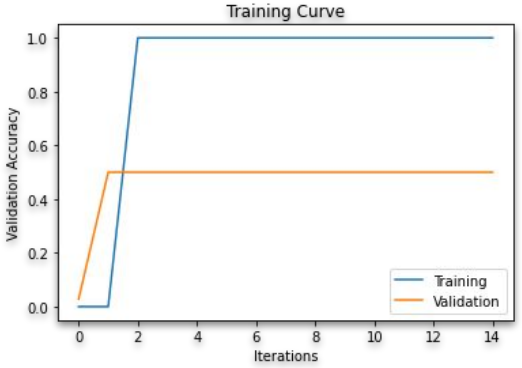


Baseline Model Comparison

ANN







Test Accuracy: 51.52%

Key Takeaways

Importance of Data Collection

How to Find the Right ML Model

Evaluate the Problem Properly

Thanks for Listening!