Using Deep Learning to Detect Pneumonia from X-Rays

HELLO!

I am Gabby Amparo

Full-Time Data Science
Student at Flatiron School

What is Pneumonia?

- According to the Mayo Clinic Website, **"Pneumonia is an infection that inflames the air sacs in one or both lungs.**
- The air sacs may fill with fluid or pus (purulent material), causing cough with phlegm or pus, fever, chills, and difficulty breathing.
- **A variety of organisms, including bacteria, viruses and fungi, can cause pneumonia.**
- Pneumonia can range in seriousness from mild to life-threatening. It is most serious for infants and young children, people older than age 65, and people with health problems or weakened immune systems."
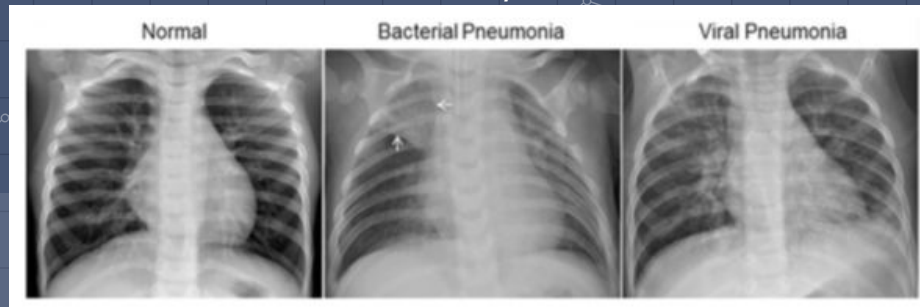
Data Sources and Background Information

- **The x-ray images used for this analysis were obtained from Guangzhou Women and Children's Medical Center, Guangzhou, China. These x-rays were obtained from pediatric patients.**
- All images that were of low quality or unreadable were removed and were reviews by three expert physicians for the purpose of AI analysis.
- The dataset was organized into 3 folders (test, training, and validation) and within each folders there were subfolders for each category (Pneumonia/Normal). In total there are 5,863 X-Ray images. The dataset was downloaded from the Kaggle website.

Programs and Libraries

The following software libraries were used within Python to conduct data analysis:

- Numpy - for mathematical computation
- Pandas - allows for data organization & analysis
- Matplotlib - for data visualization
- Seaborn - works with Matplotlib to make clean graphics
- Sklearn - machine learning
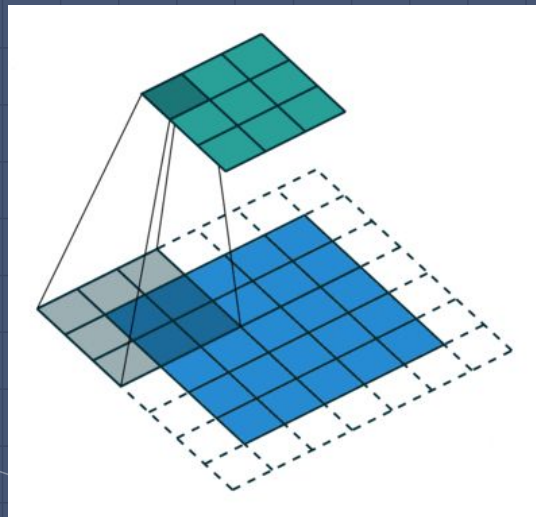- Keras & TensorFlow- deep learning and image classification
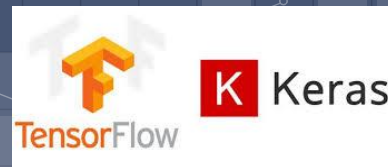
Objectives

▫ The objective of this analysis was to build a machine learning model that can detect whether or not a patient has pneumonia by classifying X-ray images.
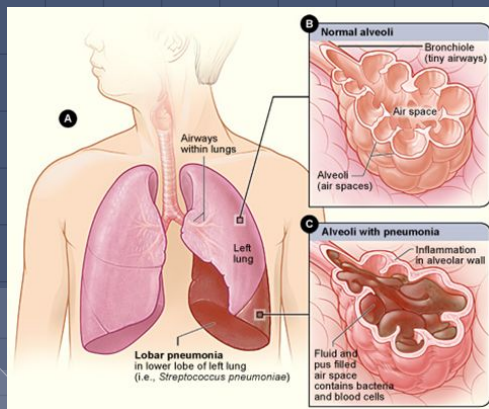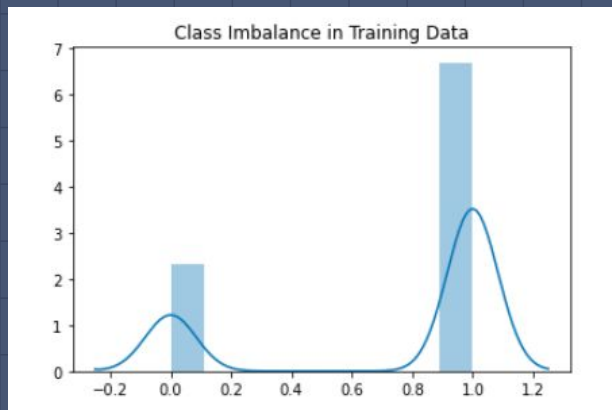
Making a CNN

- A CNN is a Convolutional Neural Network
  - Provide an alternative formulation of analyzing groups of pixels
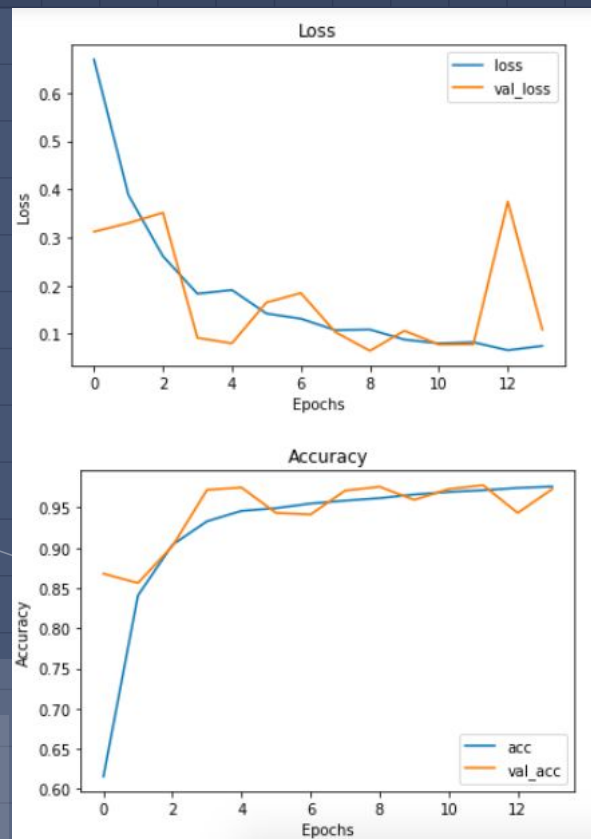
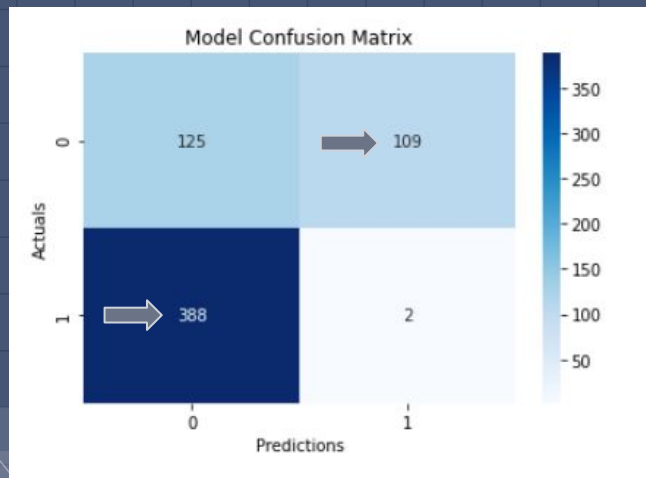# Some things to note...

Class Imbalance

Other organs are seen in the X-Ray which can confuse model

Results

Best Model: Model 4

Accuracy: 79.6%

Conclusions & Recommendations

Conclusions

- The best model had an accuracy rate of almost 80%
  - Not great but still good
- Most models had a accuracy of 70% and up
- The models most likely need more tuning
- Model is more likely to predict FN

Recommendations

- This model can be used by doctors as a template. They ultimately make the decision themselves.

Future Work

- If possible, crop out organs that are present in the X-Ray images
- I used 128 X 128 px image sizes initially, so I would like to try smaller sizes like 64 X 64 px
  - This would help greatly with model training time

THANKS!

Any questions?

You can find me at:

- gabbyamparo97@gmail.com
- @gabbyamparo