Pneumonia Detection

By: Temesgen Tesfay

Mentor: Ben Bell

Pneumonia is Lung Disease

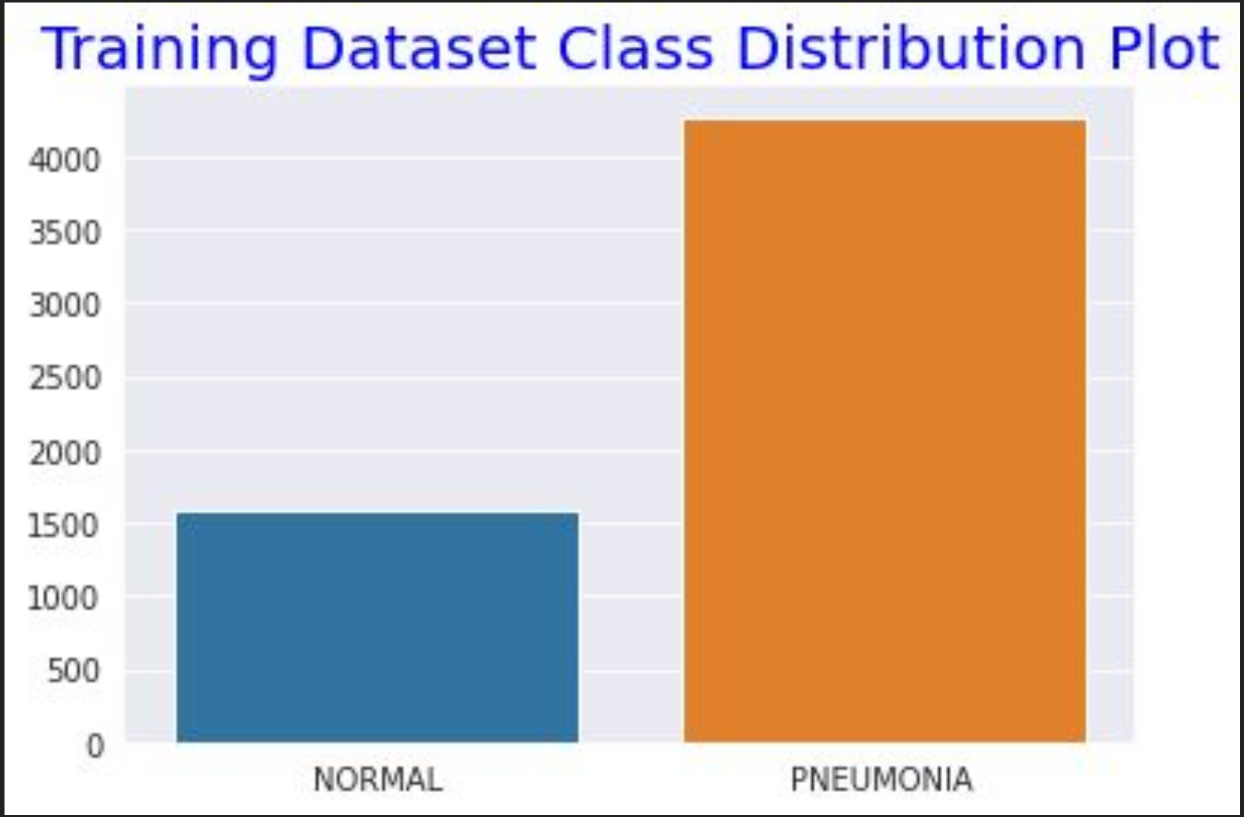
- Pneumonia is one of the top few acute diseases kills children in the world. Failure to identify the disease increases severity.
- This project aimed to improve pneumonia detection using deep learning to minimize the death rate as a result of image misclassification.

Image Classification Implemented using Convolutional Neural Network

- ❖ The project implemented deep learning using:
  - ➢ Traditional deep learning
  - ➢ Transfer learning takes pre-trained images such as ImageNet.

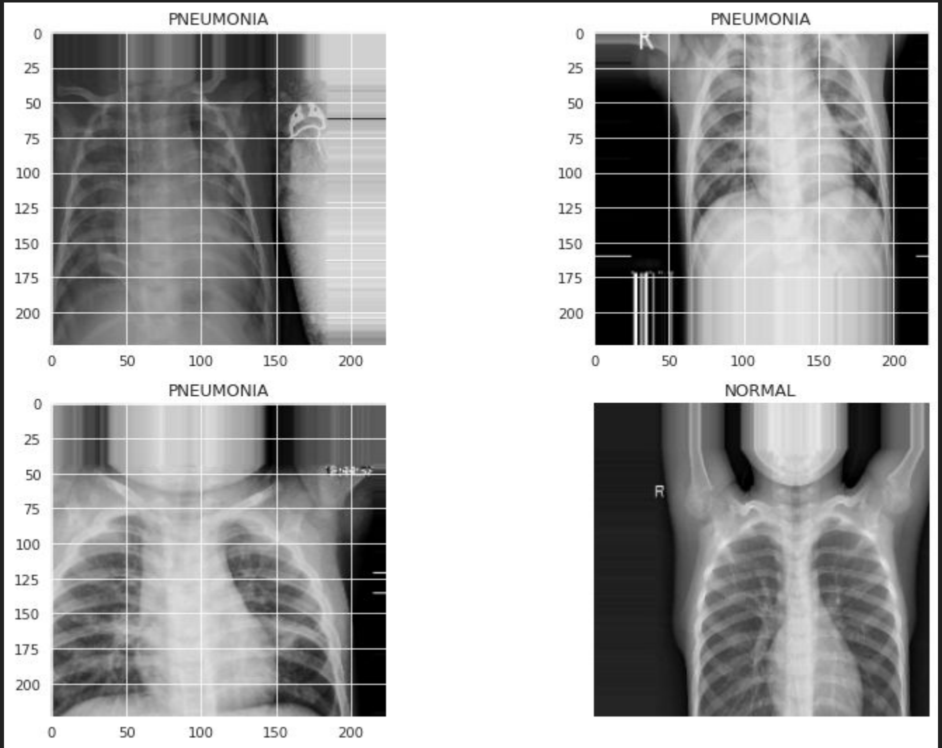
EXPLORATORY DATA ANALYSIS (EDA)

Visualization



Sample Images of Chest-X-rays

- Pneumonia-positive lungs show a wide irregular whitish area compared to the normal lung, dark brown regular pattern.



Data Augmentation (Preprocessing)

Data Augmentation is a practice to augment data artificially from the existing data when data is insufficient.

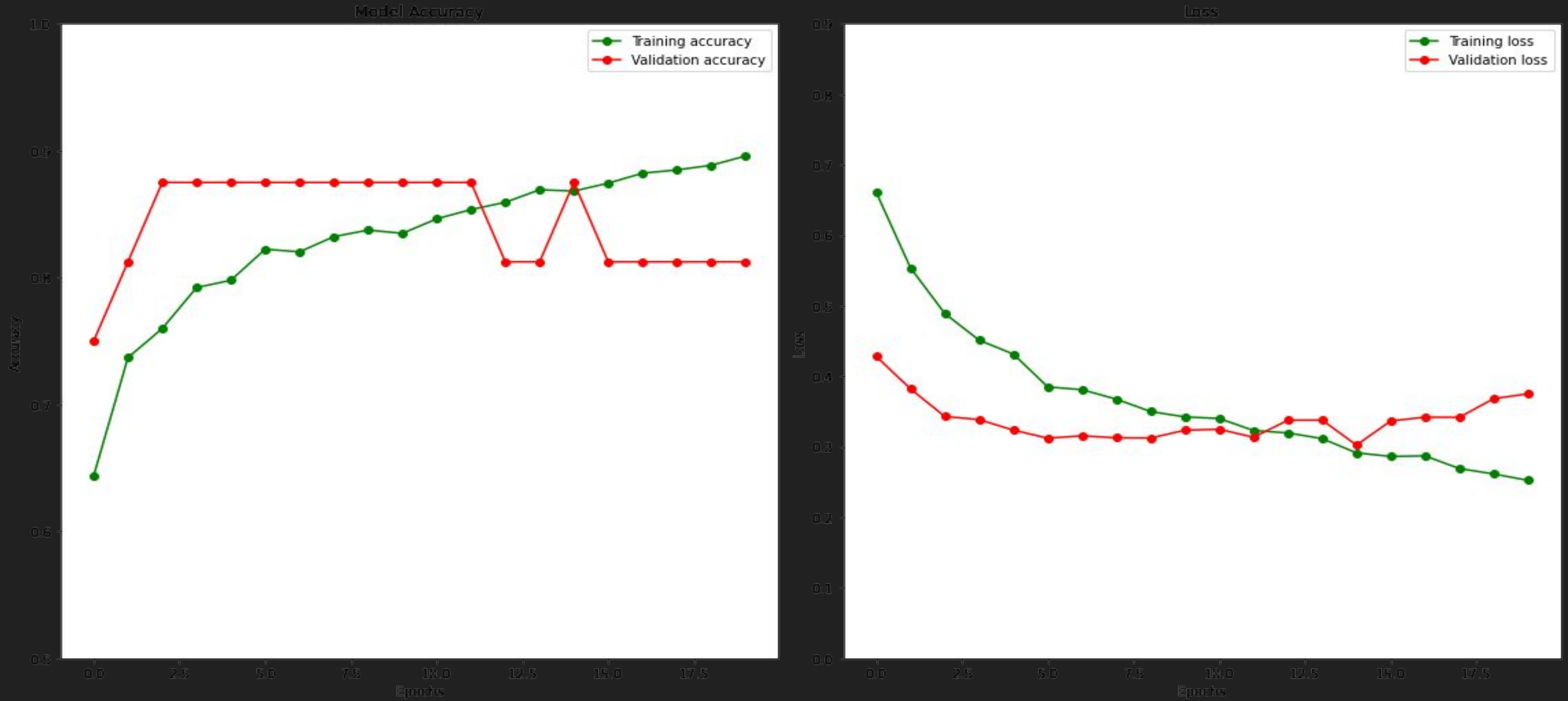
The project implemented the following data augmentation techniques

- Random-flip
- Random-rotation

Transfer Learning for Pneumonia Detection

- The ResNet50V2 model was implemented
- ResNet50V2 model is pre-trained on another dataset, but I froze all layers to preserve the features and created a new top layer which I retrained on the Pneumonia dataset to predict pneumonia.
- Transfer Learning is efficient in computational cost, and was found to have approximately the same accuracy as traditional deep learning.

Transfer Learning Accuracy and Loss Plots



Confusion Matrix at .5 Threshold

Where:

True Negative = 120

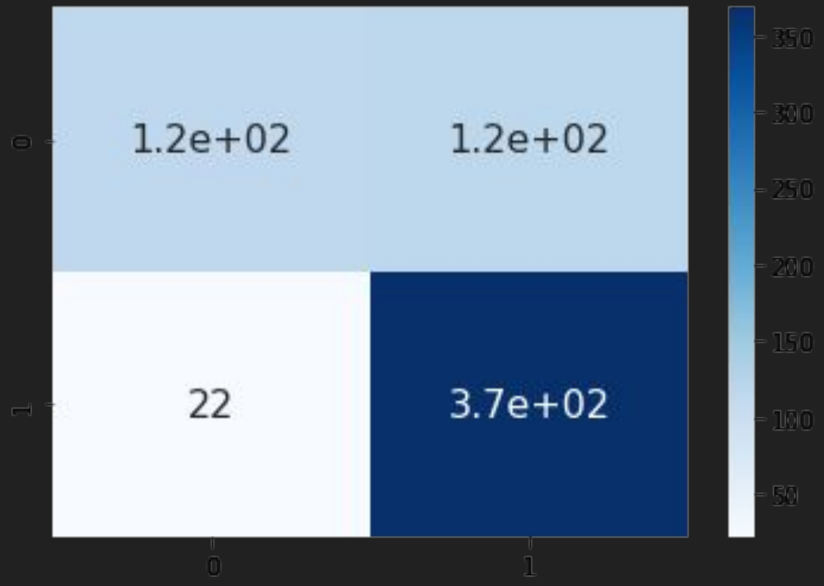
True Positive = 370

False Positive = 120

False Negative = 22

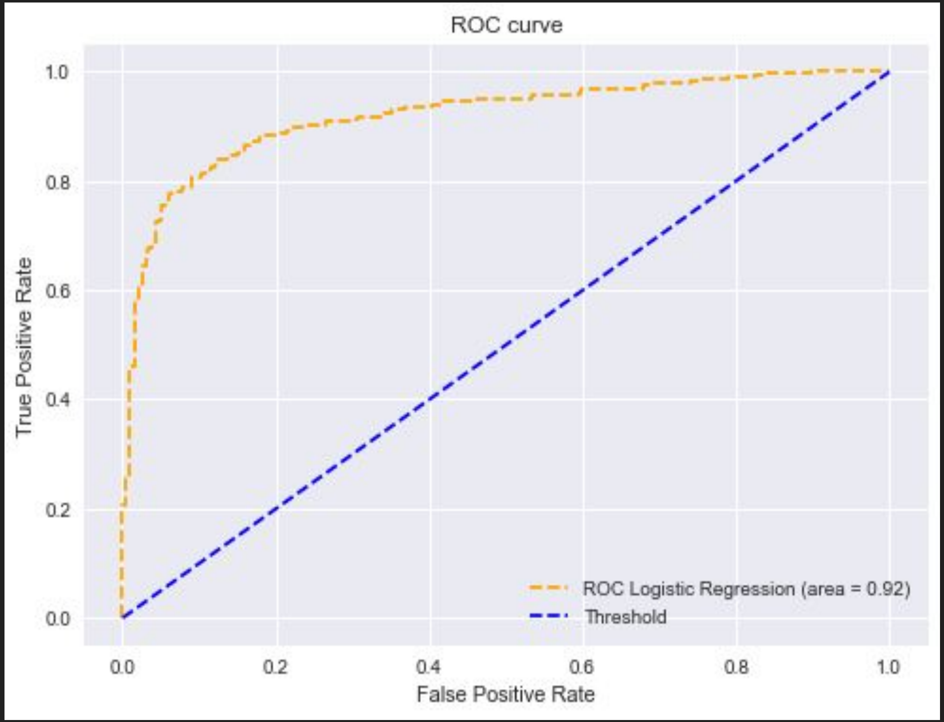
Precision = Tp/(Tp + Fp) = 0.76

Recall = Tp(Tp + Fn) = 0.94



The data imbalance between Normal and Pneumonia classes is also reflected in test data

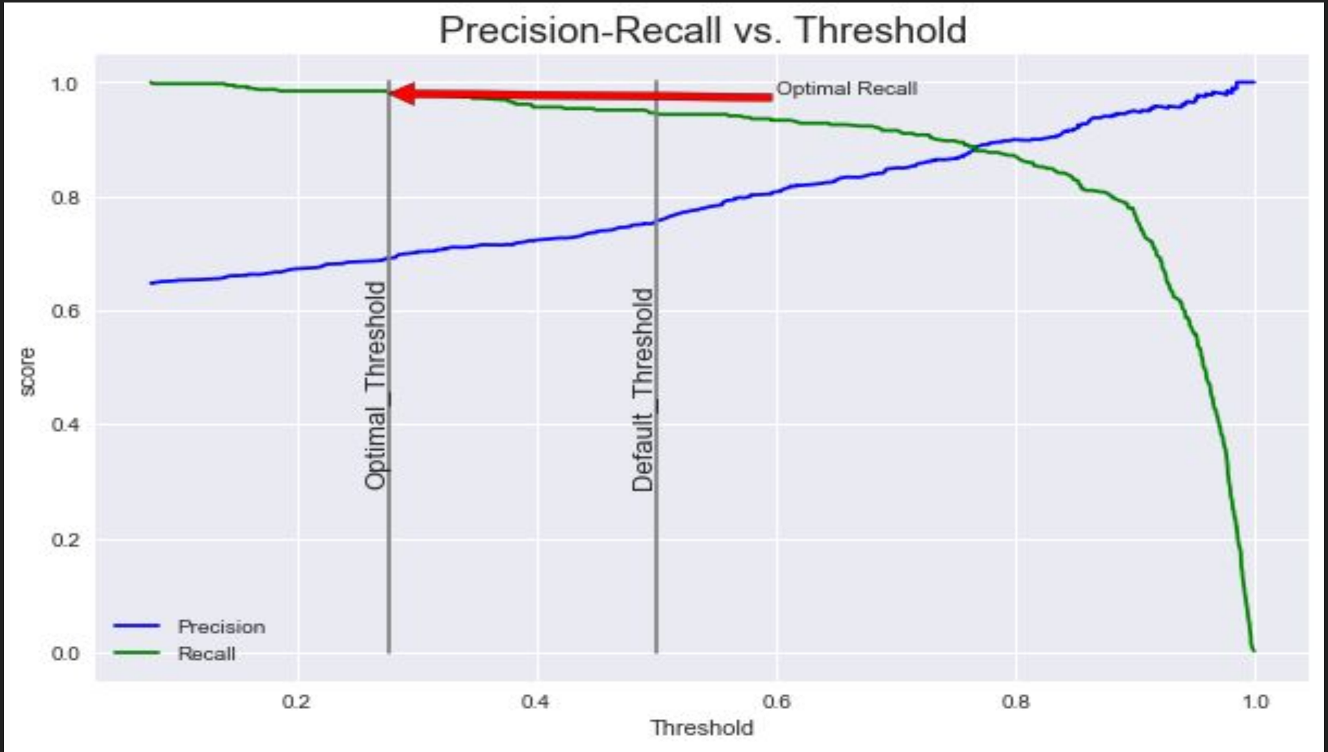
ROC Curve (Receiver Operating Characteristics Curve)



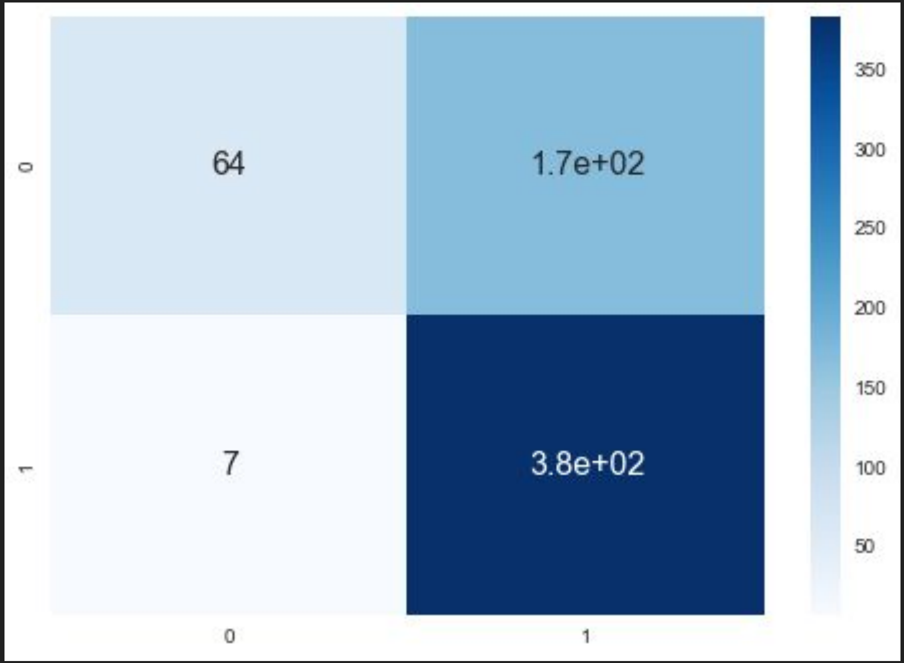
ROC Shows the trade-off between Sensitivity (TPR) and Specificity(1-FPR)

- ❖ The diagonal line is a random classifier is expected to give points lying along the diagonal where (TPR=FPR)
- ❖ AUC ~ Probabilistic Accuracy

Precision-Recall vs Threshold plot



Confusion Matrix Threshold at 0.25, optimal threshold



Recall = 98%
Precision = 69%

At threshold 0.5, model perform 94% recall and 78% precision. From optimal threshold, recall improved to 98% but precision declined to 69%.

Classification Report from 0.25 Threshold

	precision	recall	f1-score	support
0	0.90	0.27	0.42	234
1	0.69	0.98	0.81	390
accuracy			0.72	624
macro avg	0.80	0.63	0.62	624
weighted avg	0.77	0.7		

Conclusion

- Data imbalance is a critical challenge for model performance
- Transfer learning has better performance at the early stage of the epochs with a less computational cost.
- Applying data augmentation techniques minimize the insufficient data problems.