Pneumonia Detection

By: Temesgen Tesfay

Mentor: Ben Bell

Pneumonia is lung disease

- Pneumonia is one of the top few diseases that kill children in the world. The disease is curable if treated at the early stage. Failure to identify the disease increases the severity.
- This project aimed to improve the accuracy of image identification to minimize the death rate as a result of incorrect classification.

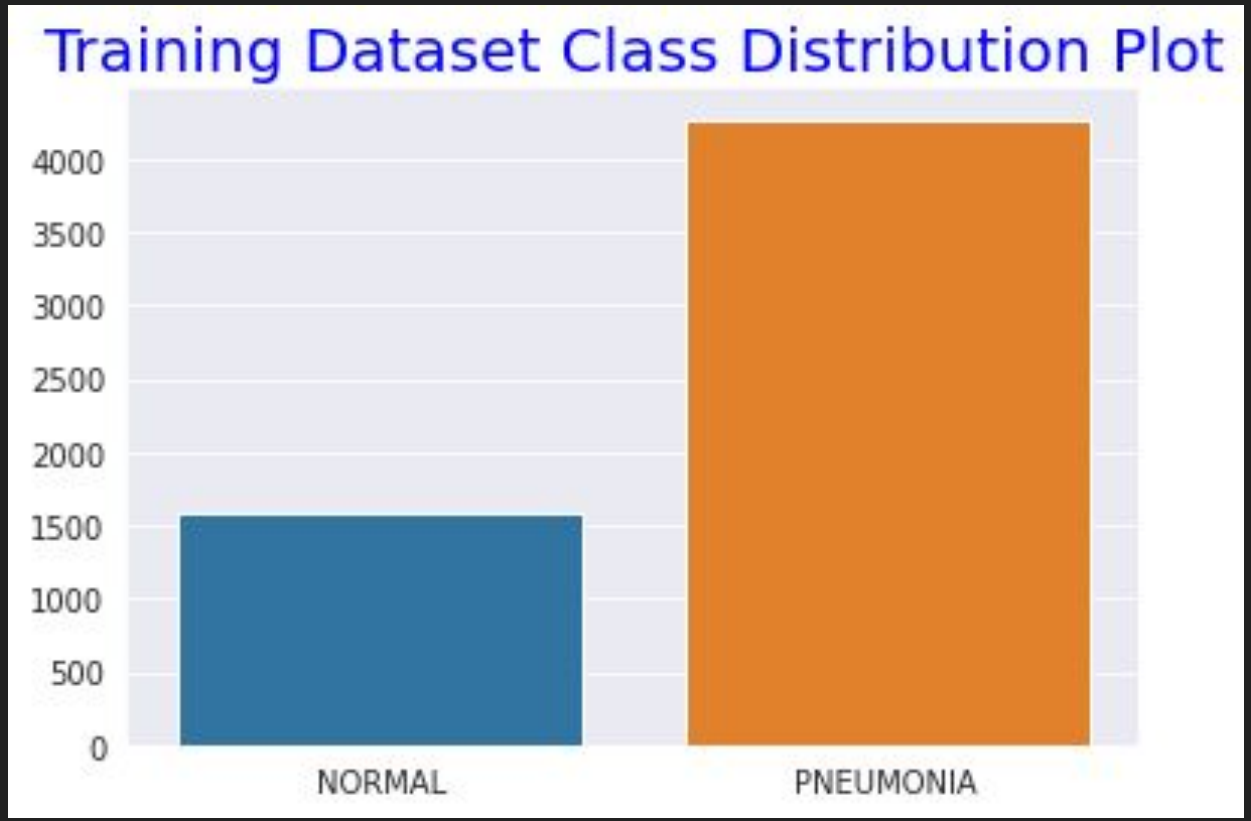
Image Classification using Convolutional Neural Network

- The project implemented deep learning using:
  - Traditional deep learning,
  - Transfer learning from pre-trained images such as imagenet

To improve model performance, multiple experimental combination carried on optimizers and regularizations functions (relu, dropout and max-pooling or average-pooling)

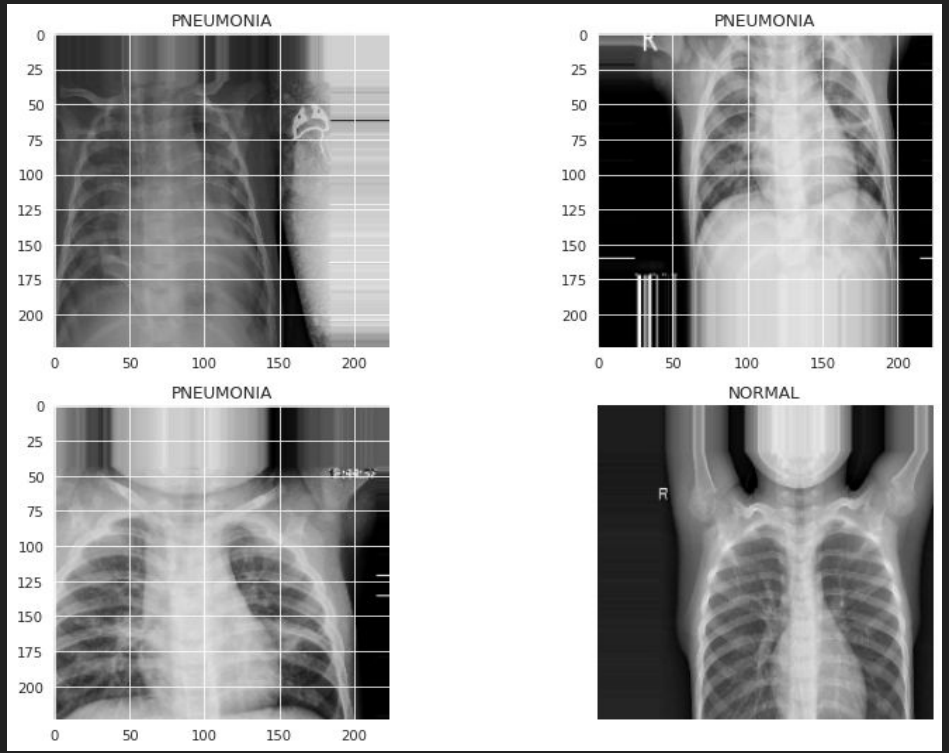
EXPLORATORY DATA ANALYSIS (EDA)

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 following data augmentation techniques are implemented :

- Random-flip,
- Random-rotation,

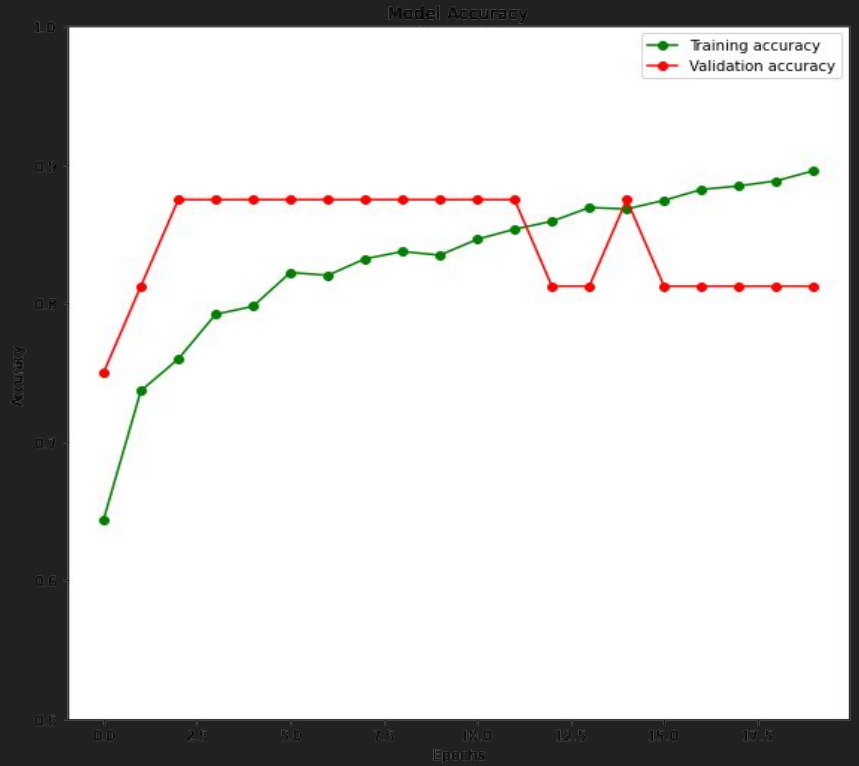
Data imbalance on the training dataset between classes lead to model overfitting.

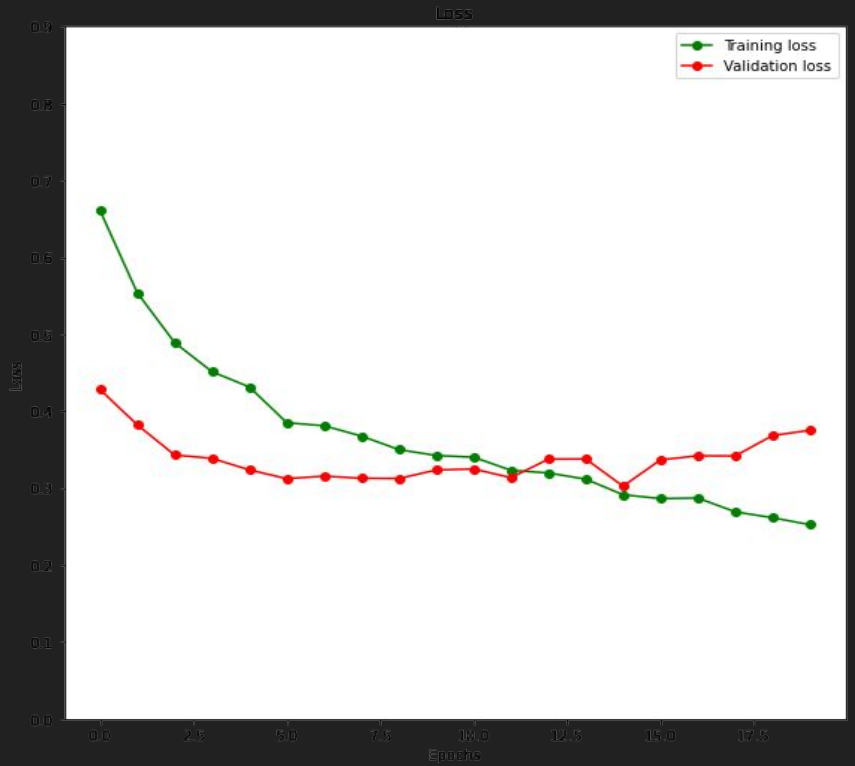
As a result of data augmentation the accuracy result improved by 3%.

Transfer Learning for Image Classification

- The model implemented ResNet50V2 keras application
- Used a base layer to connect with pre-trained images
- Transfer Learning is efficient in computational cost, and it return the accuracy approximately close to traditional deep learning. F1 score versus accuracy for transfer learning return 0.80, and 0.83 from fully connected layer.

Transfer Learning Accuracy and Loss Plots





Confusion Matrix

Where:
True Negative = 130

True Positive = 370

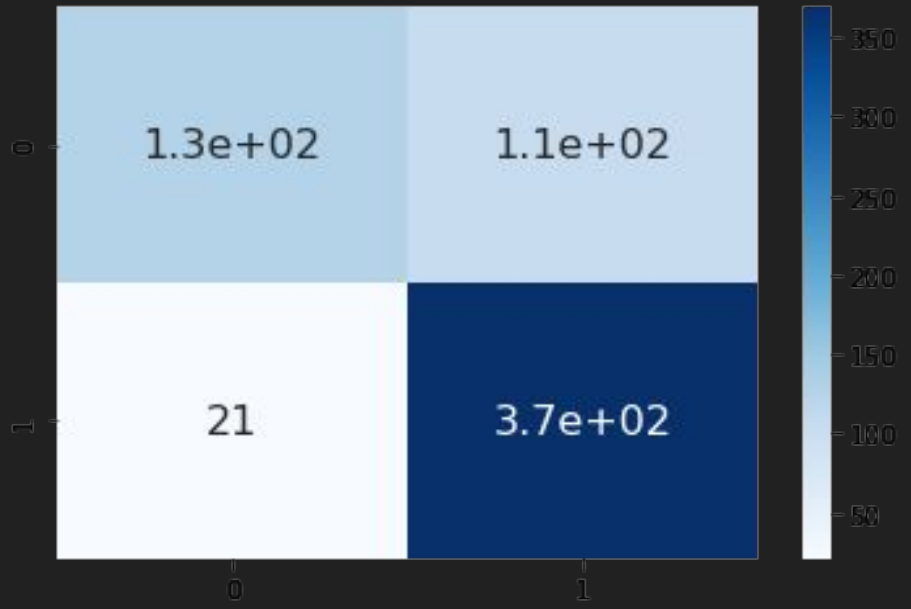
False Positive = 110

False Negative = 21

Precision = Tp/(Tp + Fp) = 0.78

Recall = Tp(Tp + Fn) = 0.95

Specificity = Tn/(Tn + fn) = 0.86



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

And accuracy is not appropriate parameter for model performance.

Model Performance Metrics from Classification_report

	Precision	Recall	F1-score	Support
0	0.86	0.55	0.67	234
1	0.78	0.95	0.85	390
accuracy			0.80	624
acro avg	0.82	0.75	0.76	624
weighted avg	0.81	0.80	0.78	624

Metric Performance Plots:- Optimal_threshold at 0.215

Conclusion

- Data imbalance a is critical challenge for model performance
- Transfer learning return better performance especially at the early stage of the epochs with less computational cost.
- Accuracy is not perform well in classification when data is imbalance in stead, F_score provide less biased outcome.
- Applying data Augmentation techniques is a sound approach to ease the insufficient data problems.