Pneumonia Image Classification

Lucas Wilkerson
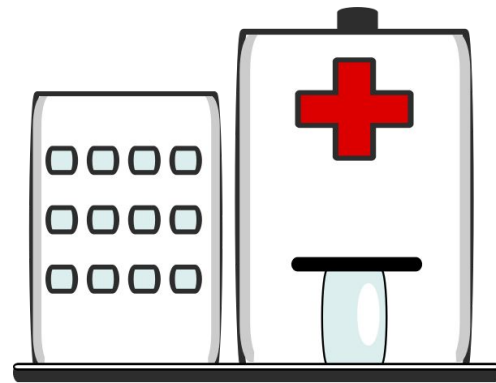
Summary

Dataset of chest X-Ray images obtained from a Kaggle dataset

Convolutional Neural Network Models Developed:

- **Goal ➜ Build a model to detect pneumonia**

Insight generated will be used to give actionable recommendations for the stakeholder (healthcare practice)
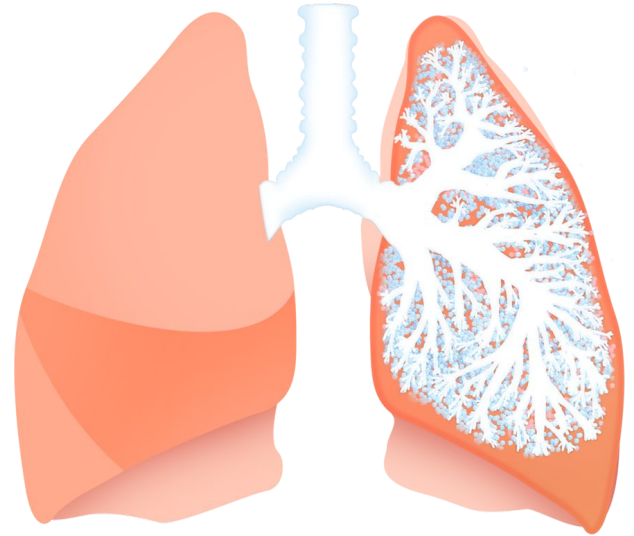
Outline

- Business Problem/ Stakeholder
- Data/Methods
- Modeling
- Conclusions/Recommendations

Business Problem/ Stakeholder

A healthcare practice group wants to improve patient outcomes by:

- Early Pneumonia Detection
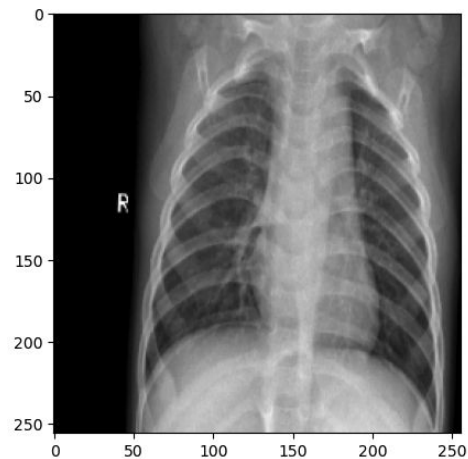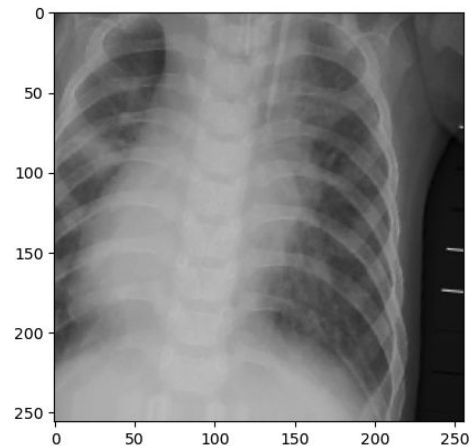- Implement Early Interventions

Data/Methods

EDA:

- 5856 Images

Three Dataset Groups:

- Train
- Test
- Validation

Data Distribution
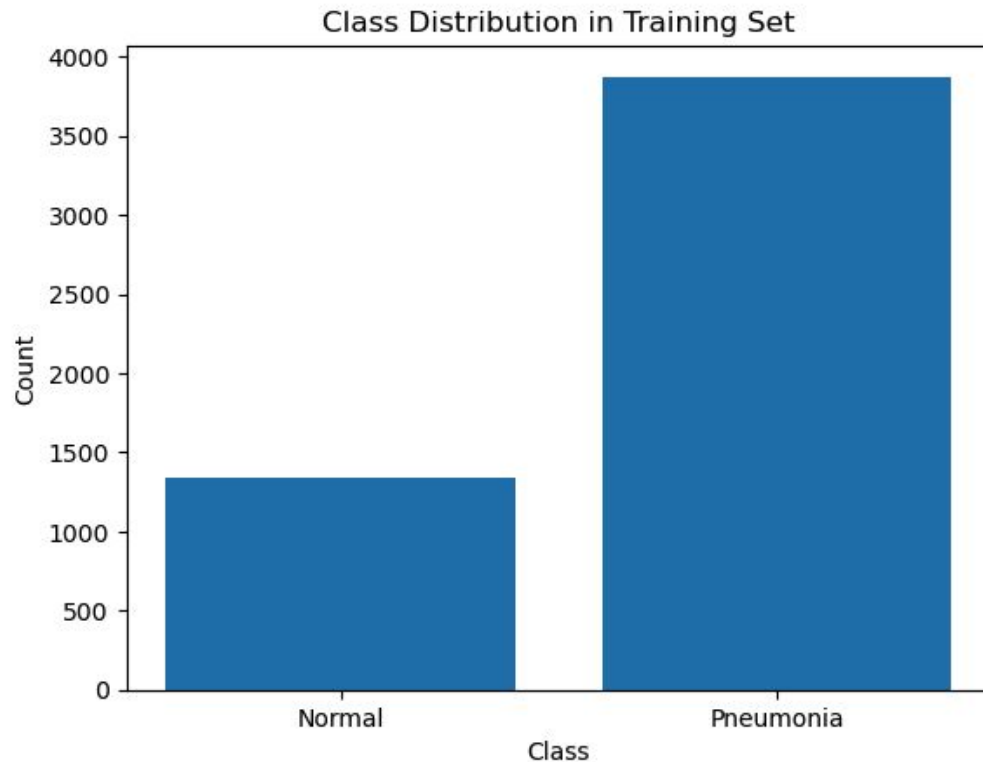
Train set: 5216 Images

- Pneumonia: 3875
- Without Pneumonia: 1341

Test set: 624 Images

- Pneumonia: 390
- Without Pneumonia: 234

Validation set: 16 images

- Pneumonia: 8
- Without Pneumonia: 8

Class Distribution in Training Set

Modeling

Best performing CNN model:

- **Model 2**

Metrics of Focus:

- **Recall Score**
- **Accuracy Score**

Model Performance: Tuned Baseline Model

Evaluation Metrics:

- Accuracy Score: **0.8221**

- Recall Score: **0.8103**

Metric	Baseline	Best Model
Train Accuracy	0.8111	0.8773
Test Accuracy	0.7500	0.8221
Train Loss	0.4179	0.2917
Test Loss	0.6010	0.5218
Train Recall	0.7510	0.8348
Test Recall	0.6513	0.8103

Conclusions/Recommendations

Early Detection:

- Early detection/ screening tool in conjunction with physician observations

Timely Treatment:

- Early interventions to prevent potentially life threatening complications

Limitations

- Higher chance of false positives with recall score as a metric
- Higher computation power needed and potentially longer runtimes depending on model complexity
- Lower sample size

Thank You!

Repository Link: https://github.com/ldwilker10/phase4_project
Email: ldwilker10@gmail.com
GitHub: @ldwilker10
LinkedIn: https://www.linkedin.com/in/lucasdukewilkerson/