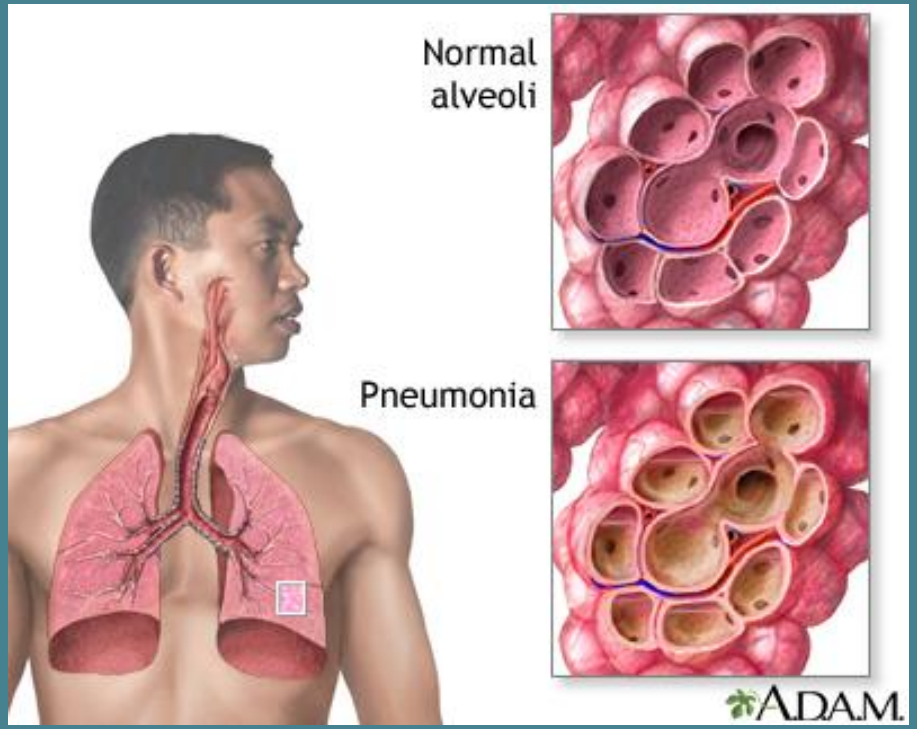
Detecting Pneumonia Using Image Processing

Nichole Hartz

Capstone 3

Pneumonia



- Lung infection that inflames air sacs
- Air sacs fill with liquid or pus
- Range from mild to severe
- Bacterial, viral or fungal

Diagnosing

- Diagnosed in variety of ways
- Commonly with chest X-ray
- Find infection & severity of spread
- Look for white spots in lungs
- Interpreted by radiologist

The Problem:

Increase in chest X-ray volumes

Significant backlogs

Average wait time: 30 days

How can HUP's Lung Center implement a system to correctly classify chest X-rays as either pneumonia detected or normal with at least an 80% precision score by the end of 2021?

Data Information

Guangzhou Women and Children's Medical Center

5,863 X-ray images (anterior-posterior)

File format: JPEG

Pediatric patients (1 to 5 years old)

Pre-split into training, testing and validation folders

Prescreen for quality control

Graded by expert physicians

Image 1 from Training Data

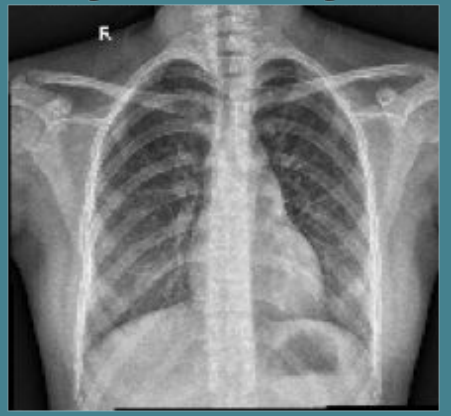


Image 5216 from Training Data

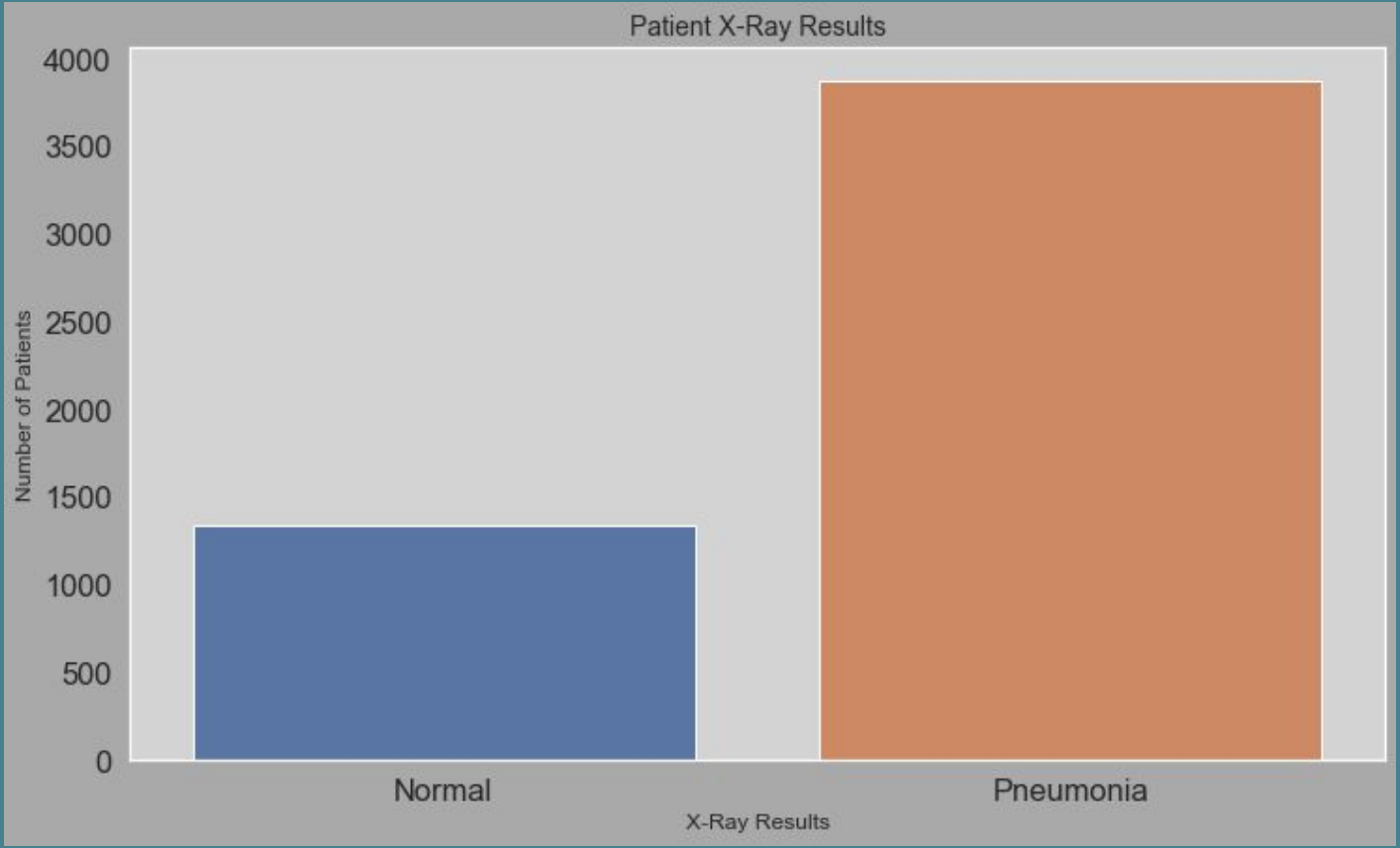
Labels

Binary categories

Class imbalance

Normal: 26%

Pneumonia: 74%



Model Overview

Supervised Learning: Images Labelled

Binary Classification: Pneumonia or Normal

Class Imbalance: More Pneumonia than Normal

Tools: Keras

Modeling Steps

Preprocess

- Shuffle
- Equalize
- Reshape
- Weight classes
  - Normal: 1.94
  - Pneumonia: 0.67

Augment Data

- Re-scale
- Randomly
  - Rotate 25°
  - Zoom 20%
  - Shift width 10%
  - Shift height 10%
- Flip horizontally

Build Model

- Four Convolutions
  - Traditional
  - Separable
- Max Pooling
- Batch Normalization
- Flatten
- Dropout 30%
- Two Dense
  - Relu
  - Sigmoid

Fit Model

- Optimizer: ADAM
- Loss: binary cross entropy
- Metrics: accuracy, F1, precision, recall
- Epochs: 6
- Validation data
- Class weights
- Learning rate reduction

Evaluate

- Evaluate on test set

Model Performance

Training Data

- Loss: 0.15
- Accuary: 0.94
- F1: 0.96
- Precision: 0.98
- Recall: 0.94

Validation Data

- Loss: 1.07
- Accuary: 0.63
- F1: 0.72
- Precision: 0.57
- Recall: 1.00

Testing Data

- Loss: 0.25
- Accuary: 0.92
- F1: 0.93
- Precision: 0.91
- Recall: 0.95

Model is overfit

Constraints

- Small dataset for image processing
- Pre-split data
  - Small number of images for validation
- Not labelled for severity of infection
- Not labelled by cause of infection
- Only looking at pediatric patients

Ideas to Improve Model Performance

- Optimize modeling layers
- Gather adult patient data
- Relabel images
  - Severity
  - Type of infection
- Change split ratio

Further Research

- Detect infection severity
- Classify more illnesses
- Research areas of infections

Summary

- Model: Keras convolution neural network
- Model overfit
  - Test precision: 0.91
  - Val precision: 0.57
- Constraints
  - Small dataset
  - Small validation set
  - Data not labelled for severity
  - Only pediatric X-rays
- Improvements
  - Rework layers
  - Gather adult data
  - Relabel for severity
  - Relabel for cause of infection
  - Resplit data
- Further research
  - Disease severity
  - Detect different disease types
  - Link infection area to disease severity