Predicting Lung and Heart Disease From Chest X-Rays

Angad Pfuetzner

The Problem

Can we automate disease detection from chest X-rays to save radiologists countless hours of tedious inspection?
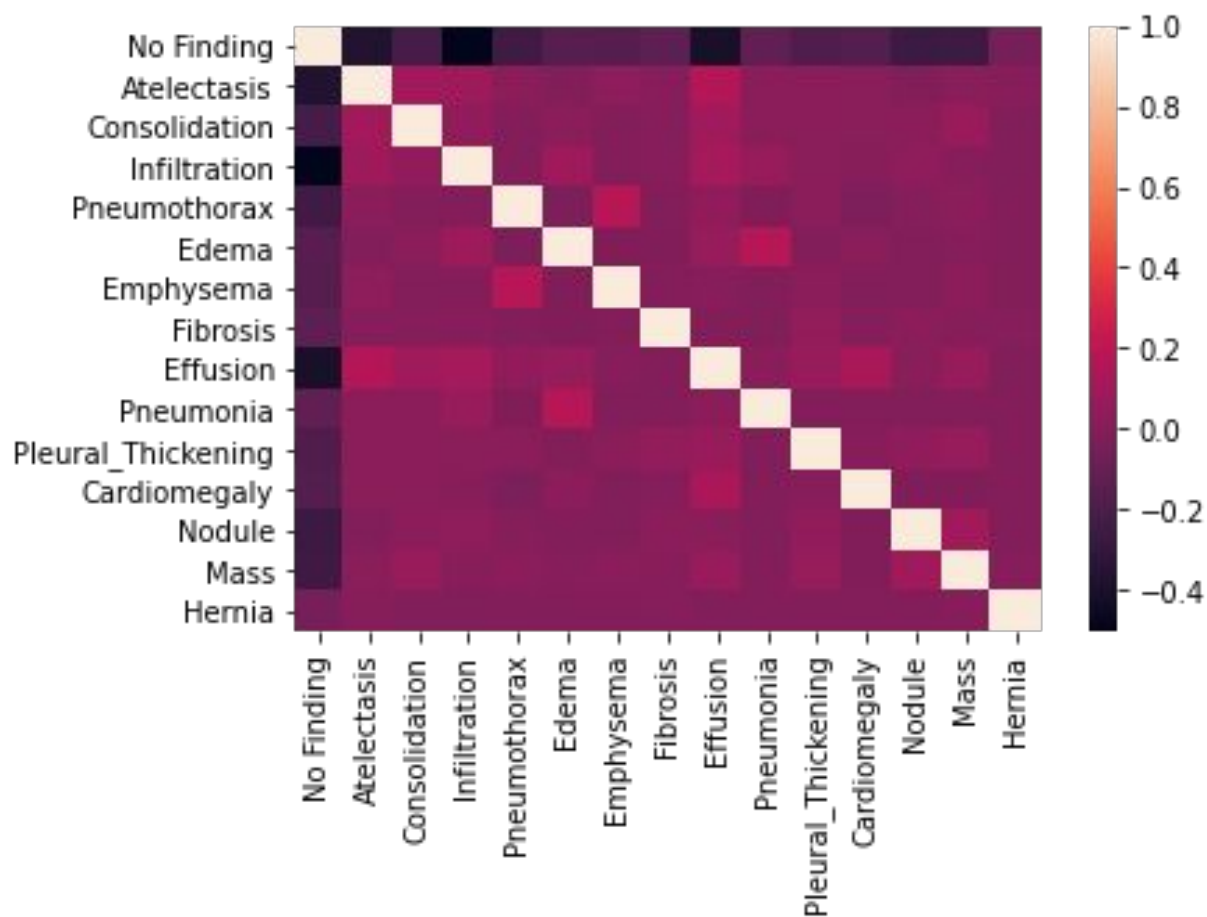
Who Cares?

Doctors

Patients

Data Sources

Data Exploration

- 112 120 X-ray images
- 30 805 unique patients
- Conditions: Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural Thickening, Cardiomegaly, Nodule, Mass, Hernia

Convert to a binary classification problem

Label	# of Images
No Finding	60 361
Atelectasis	11 559
Consolidation	4 667
Infiltration	19 894
Pneumothorax	5302
Edema	2303
Emphysema	2516
Fibrosis	1 686
Effusion	13 317
Pneumonia	1 431
Pleural Thickening	3 385
Cardiomegaly	2 776
Nodule	6 331
Mass	5 782
Hernia	227

Preprocessing
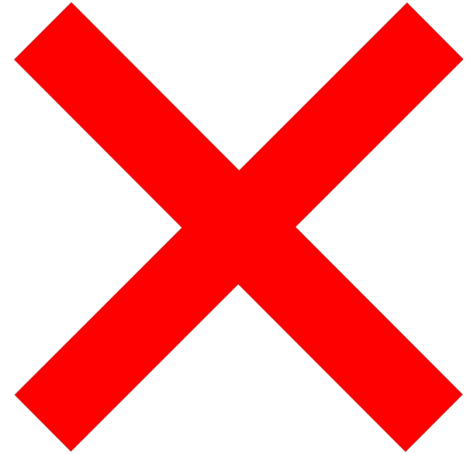

Original Image   Histogram Equalization   Adaptive Histogram Equalization

Preprocessing

Modeling

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	320
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling2	(None, 62, 62, 64)	0
conv2d_2 (Conv2D)	(None, 60, 60, 128)	73856
max_pooling2d_2 (MaxPooling2	(None, 30, 30, 128)	0
conv2d_3 (Conv2D)	(None, 28, 28, 128)	147584
max_pooling2d_3 (MaxPooling2	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense_4 (Dense)	(None, 256)	6422784
dense_5 (Dense)	(None, 2)	514

Total params: 6,663,554
Trainable params: 6,663,554
Non-trainable params: 0

Results

Model Prediction

[0.5262185 0.4737815]
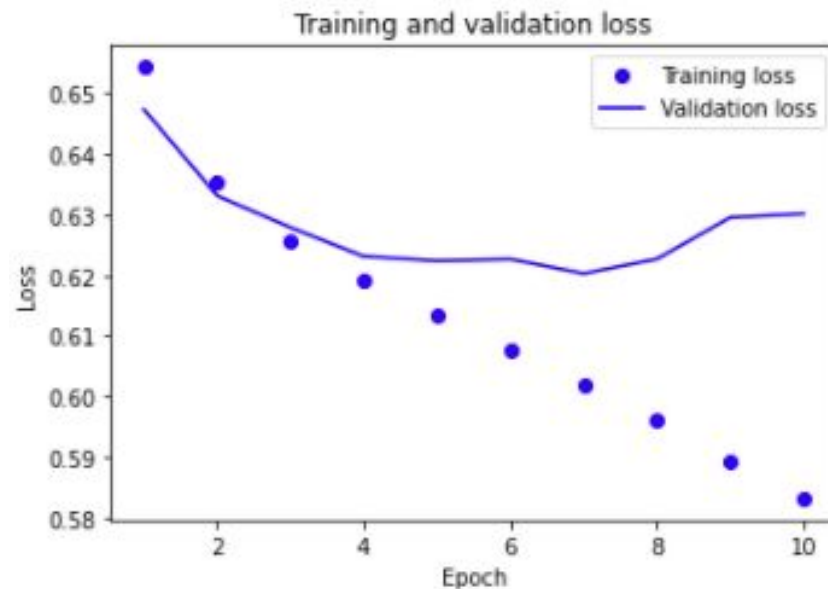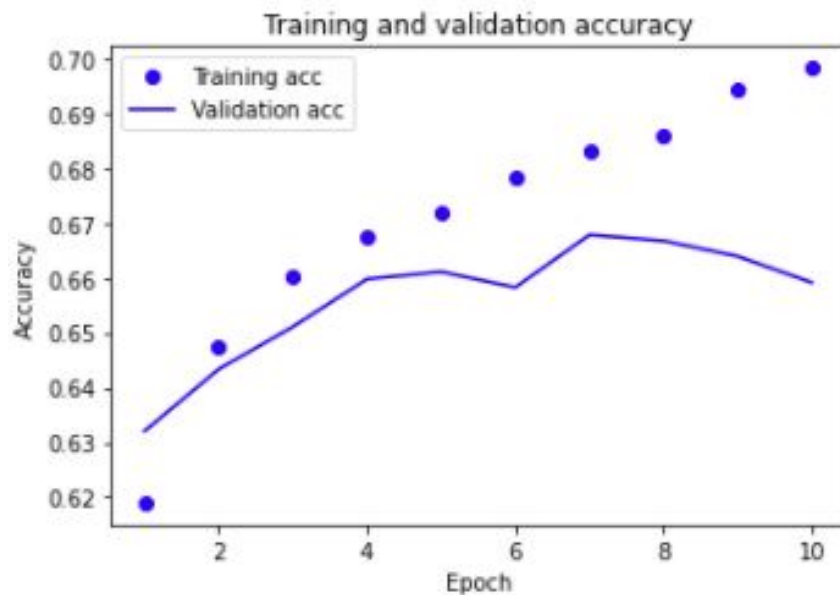
Final Accuracy: 67%

Further Study

- Additional preprocessing strategies
  - Ex. Contrast, Brightness
- Train the model with a mix of image types
  - Rotated or randomly preprocessed images to increase model robustness
- Increased model complexity

Thank You!

Github: https://github.com/Angad99/Chest-X-Ray-Disease-Detection
LinkedIn: https://www.linkedin.com/in/angadpfuetzner/
Email: angad.pfuetzner@mail.mcgill.ca