Detecting Cardiomegaly from Chest X-Rays

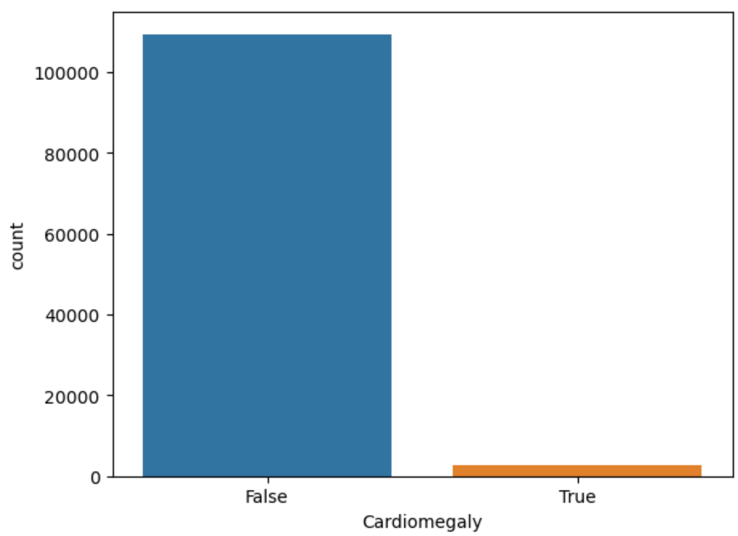
Introduction

Cardiomegaly, the condition of enlargement of the heart, is diagnosed through medical imaging. X-rays can be used to determine if the cardiothoracic ratio (the ratio of the maximal cardiac width to the maximal thoracic width) is greater than 50%. This quantifiable enlargement comes with distinct patterns on the image, making convolutional neural nets promising for cardiomegaly detection from images.

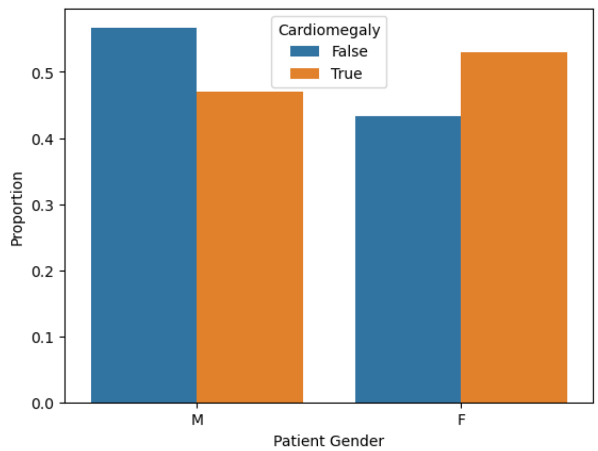
Using chest X-rays published by the National Institutes of Health labeled with natural language processing, I created a CNN model to predict cardiomegaly from X-ray images with an accuracy of 77.75%, precision of 79.85%, and recall of 74.23%. This process can be repeated for other diseases diagnosed by medical imaging such as pneumonia.

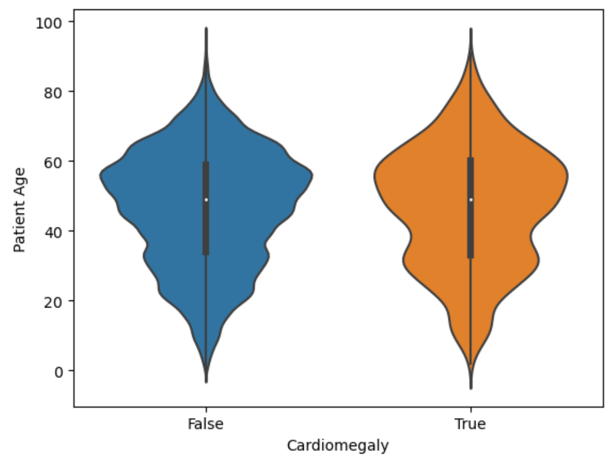
Data Wrangling

The data consisted of a csv file including the image names for 112,120 X-ray images, findings, and patient demographics. I created a target column by determining the findings that contain the string “Cardiomegaly”. 2776 images contained patients with cardiomegaly, representing 2.5% of the full dataset.

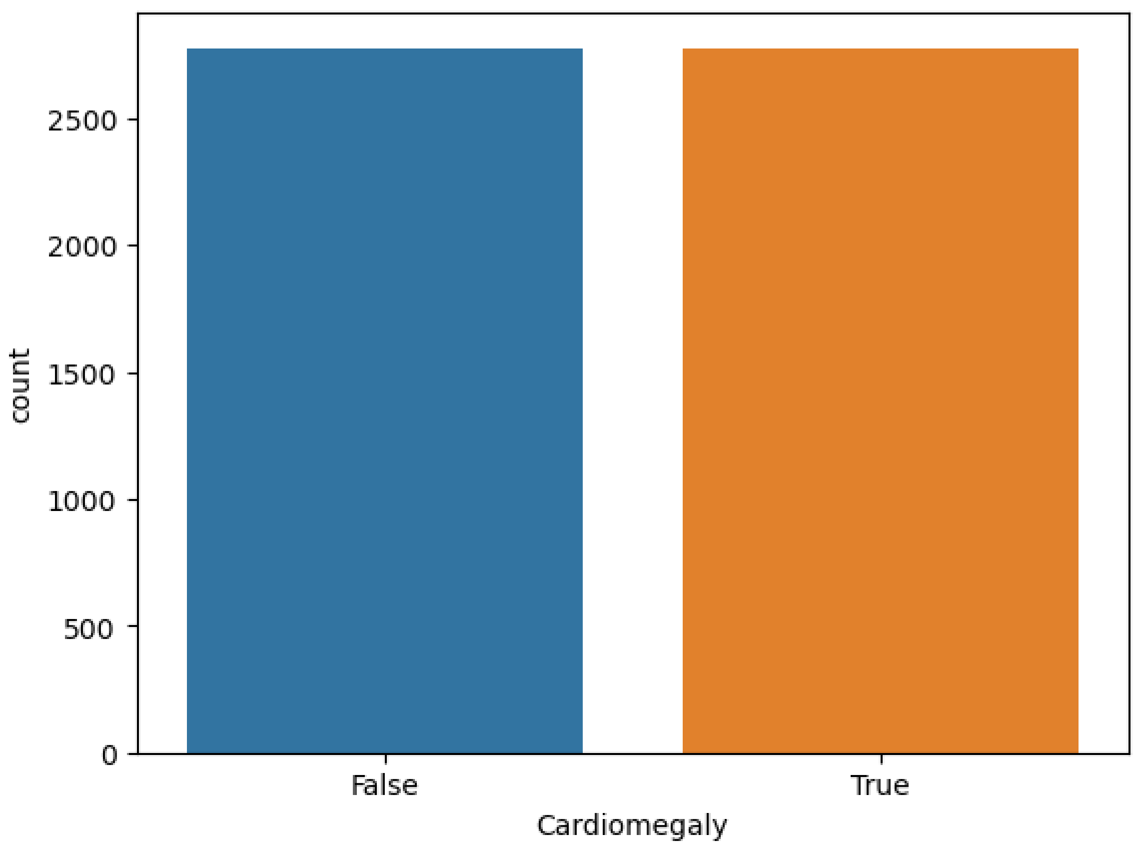


Of the patients with cardiomegaly, the age distributions were nearly identical but there is a higher proportion of female patients than male.

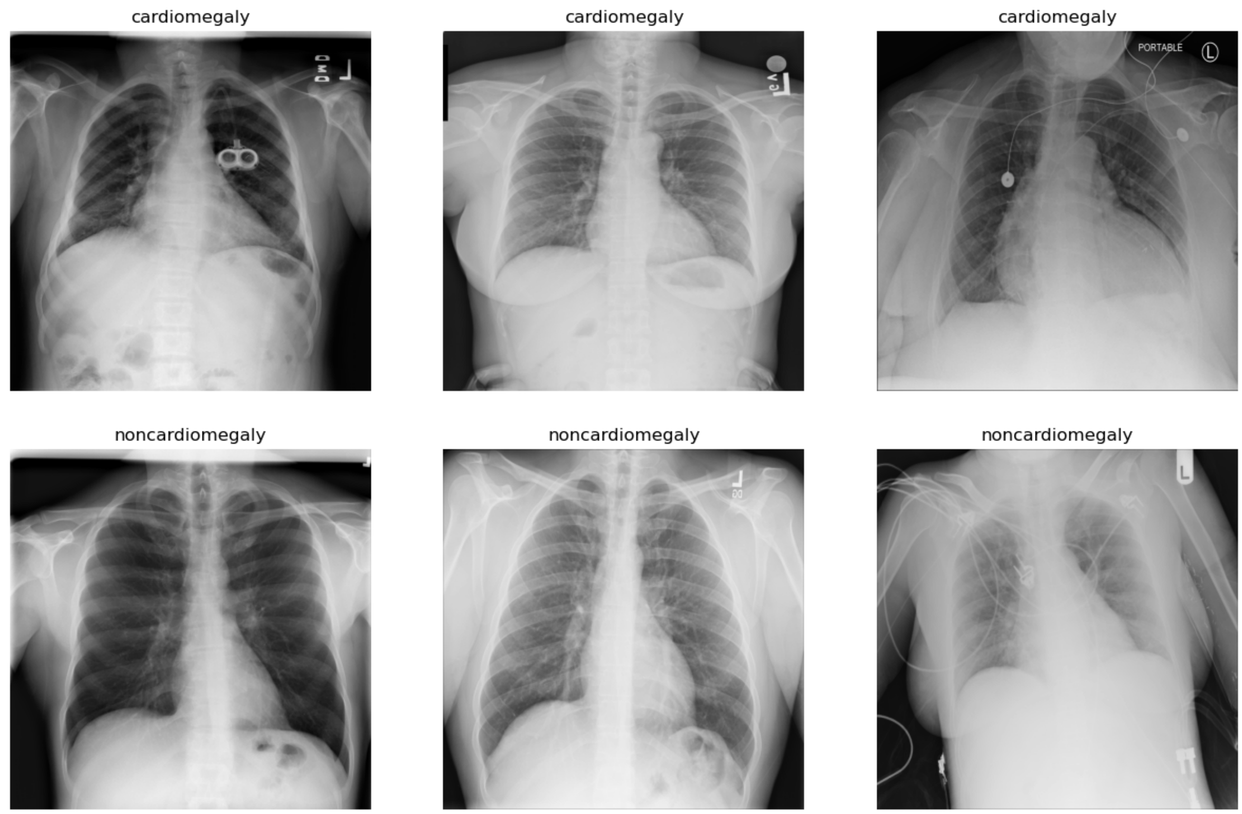




In order to balance the classes as well as decrease computation time for the scope of this project, I undersampled the data, taking 2776 of each class to make an even split.



The X-ray images are sized 1024 by 1024 pixels and are grayscale.



To prepare the data for training, validation and testing, I created directories that included 60% of the images for training, 20% for validation, and 20% for testing. Additionally, I created a subset of just 1000 images for faster tuning.

Image Pre-processing

I preprocessed images with the resnet50 preprocess_input function, which zero-centered each color channel with respect to the ImageNet dataset but did not scale the data.

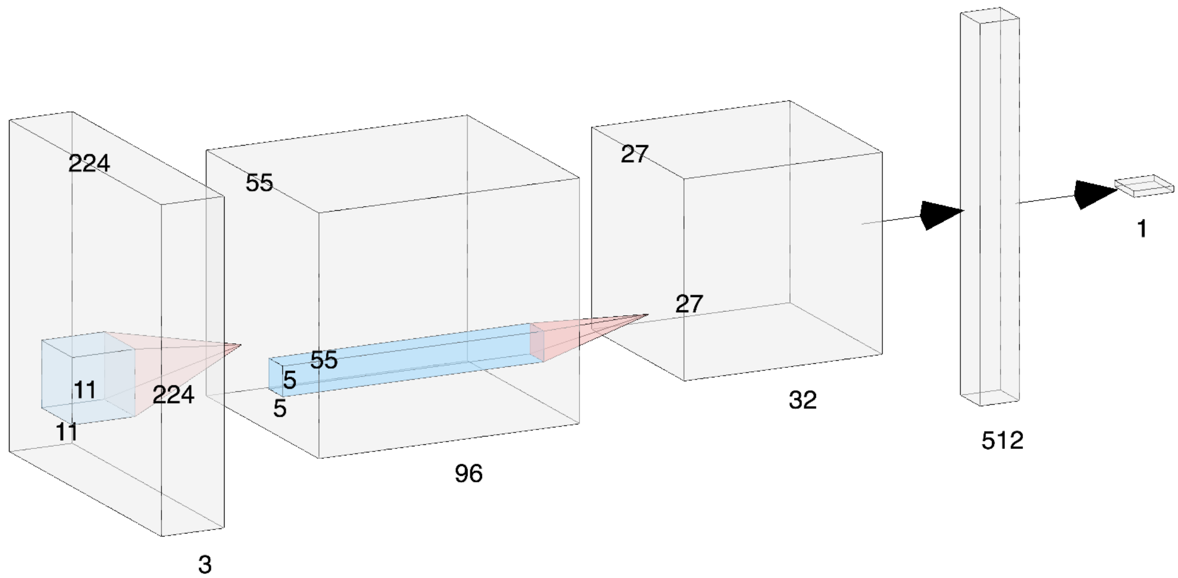
Model Architecture

I tried the following models:

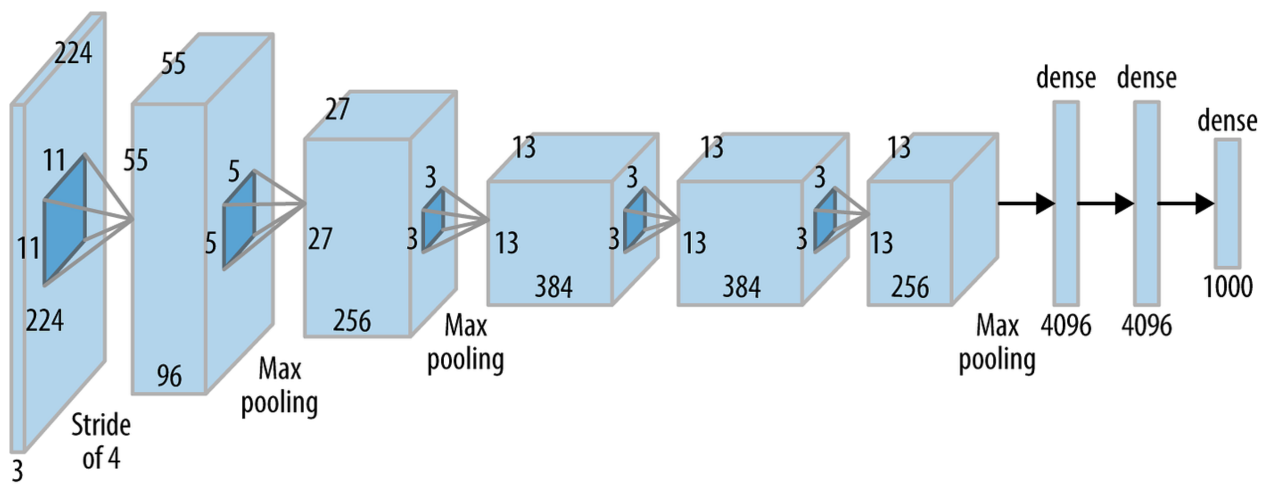
- A simple CNN
- AlexNet
- ResNet50
- An adapted ResNet50

My simple CNN architecture consists of the following:

- Convolutional layer with 96 filters of size 11x11 with a stride of 3, activated by relu
- Max pooling layer
- Convolutional layer with 32 filters of size 5x5, activated by relu
- Max pooling layer
- Dense layer with 512 neurons activated by relu
- Output layer with 1 neuron activated by sigmoid function

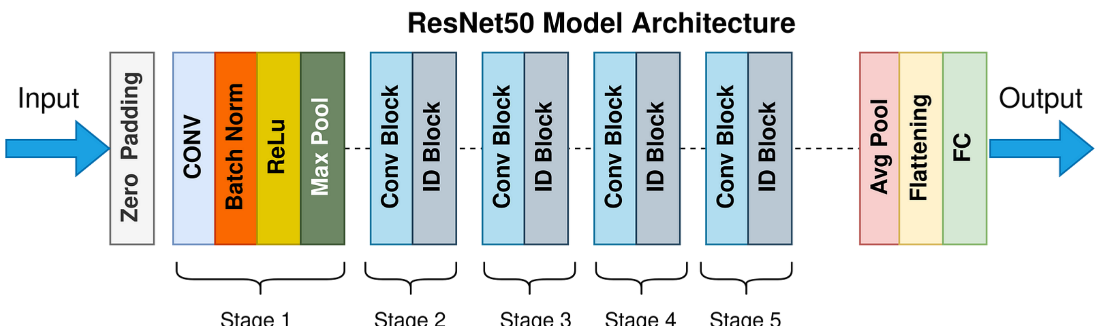


Here is the classic AlexNet architecture. I replaced the final dense layer with one with one neuron activated by sigmoid to match my binary classification problem.



Alexnet Block Diagram (source:oreilly.com)

The ResNet50 architecture consists of five stages of convolution, batch normalization, and max pooling before average pooling into a fully connected layer.

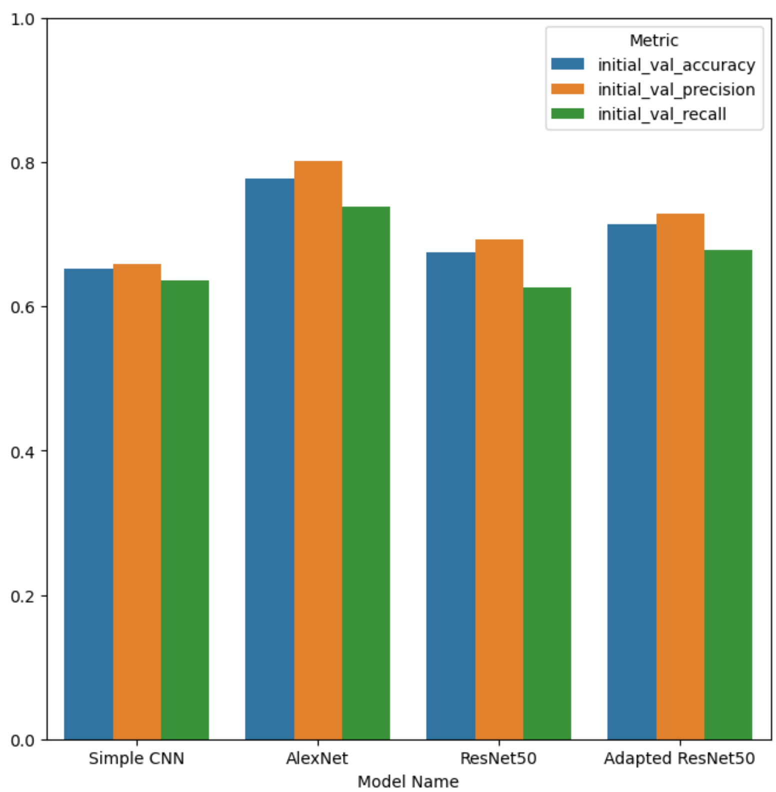


[Wikimedia commons]

In my adapted ResNet50 architecture, an additional dense layer with 256 neurons occurs before the output layer.

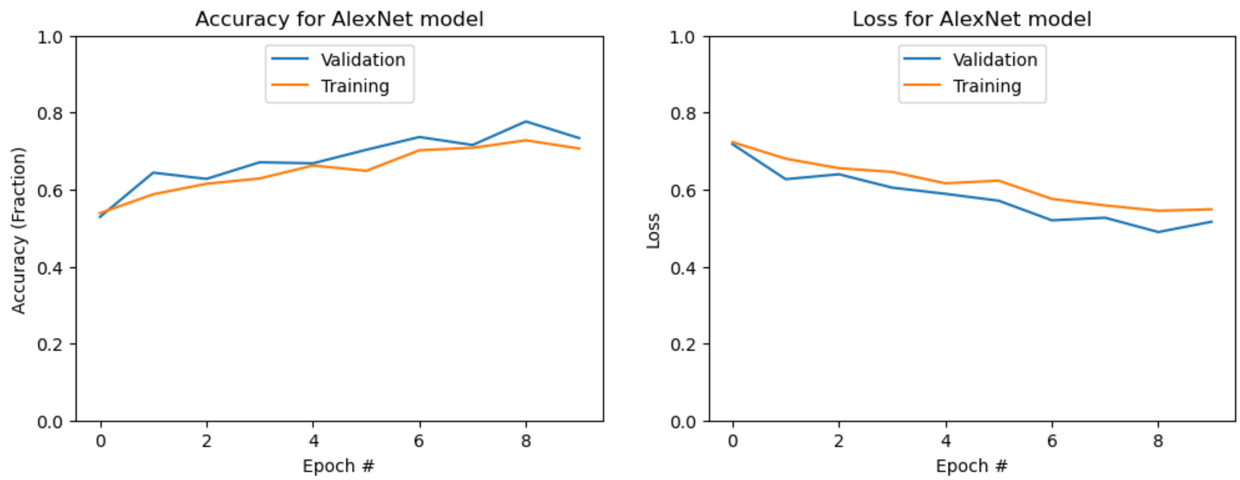
Model Selection

I began by training the four models on the full dataset for 10 epochs with a learning rate of 0.0001.



For all models, recall was the lowest metric and precision the highest. The highest performing model was AlexNet and the lowest was the Simple CNN. The

AlexNet obtained a validation accuracy of 0.7770 and validation recall of 0.7374. Additionally, the loss for the model had yet to fully even out, leading me to believe that a higher number of epochs may be beneficial. However, the overall validation loss remains quite large at 0.4896. Additionally, the Adapted ResNet50 model outperformed the original ResNet50, possibly as there are more trainable weights in the fully connected layers.

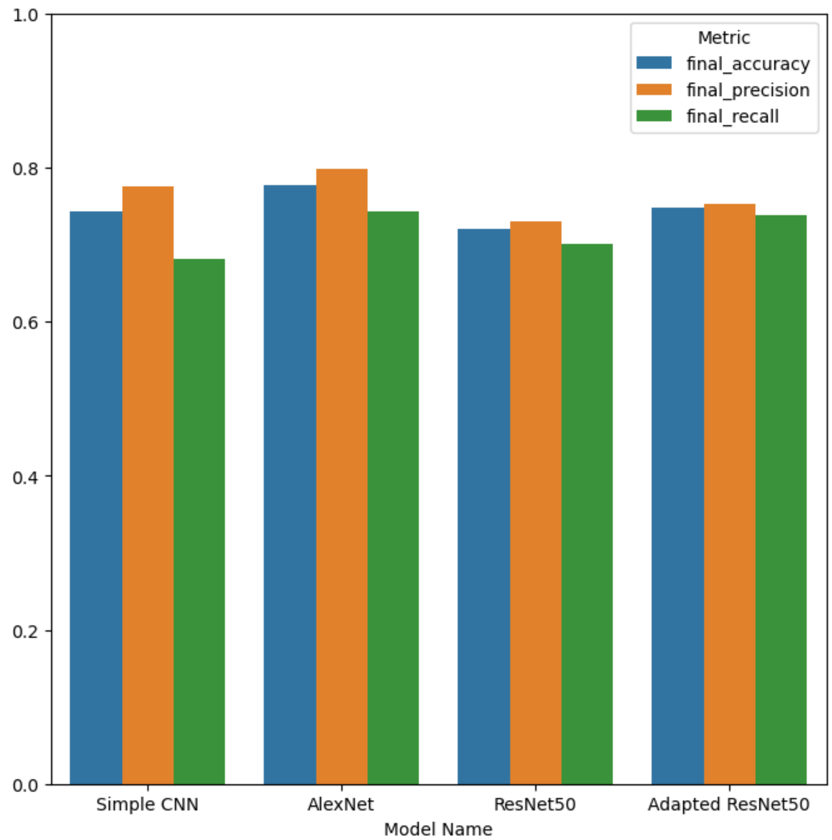


I then tuned the four models to find the optimal learning rate and batch size, which will affect the batch normalization in the AlexNet and ResNet models. I used the partial dataset of 1000 images to explore this.

Model	Optimal Learning Rate	Optimal Batch Size
Simple CNN	1e-4	45
AlexNet	1e-5	15
ResNet50	1e-5	15
Adapted ResNet50	1e-4	45

A larger learning rate appears to be associated with a larger optimal batch size. However, no learning rates smaller than 1e-6 were optimal for any model.

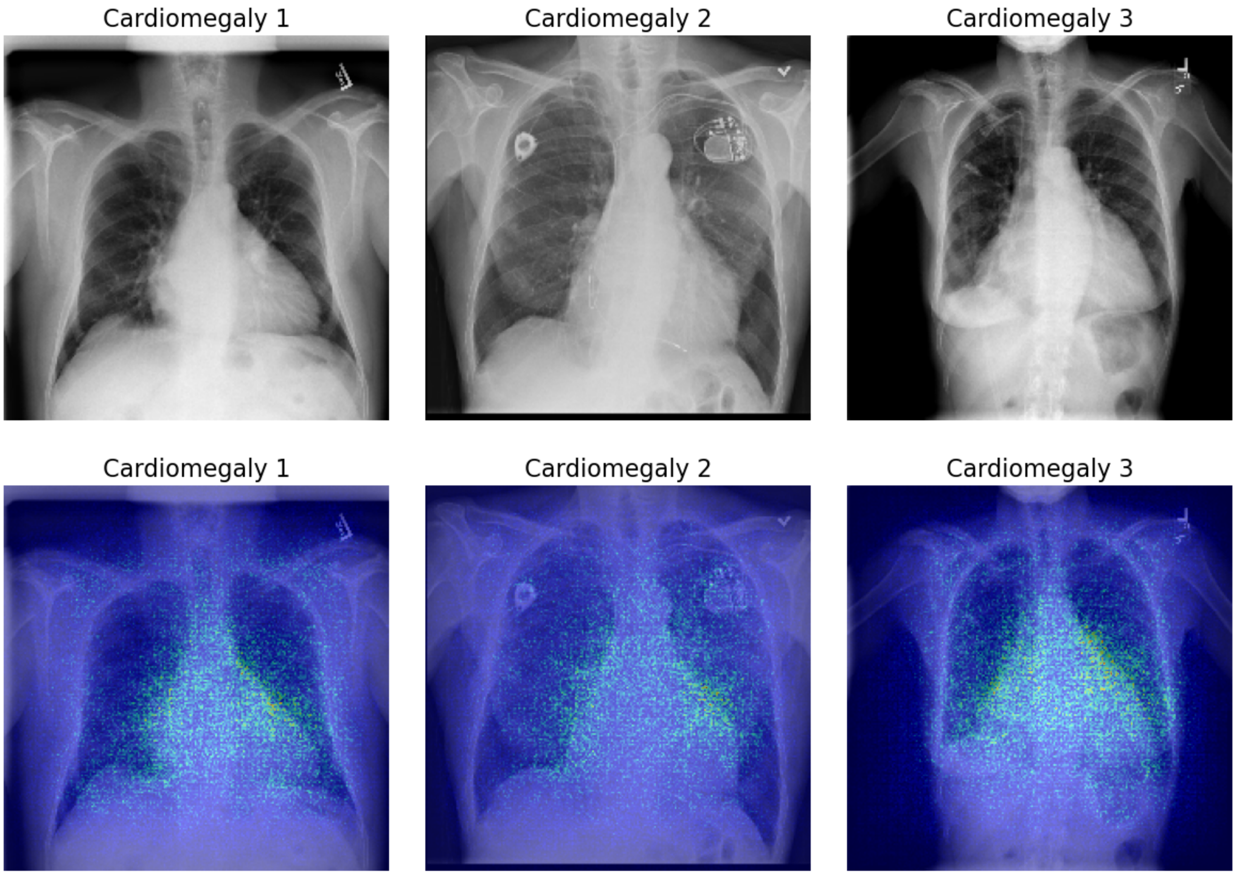
In preparing the final models, I trained them on the full dataset for 40 epochs but with an early stopper for when the validation loss had not improved in 5 epochs. Using a model checkpoint, I loaded the weights associated with the smallest validation loss and evaluated the models using the testing dataset.



As with the initial results, the AlexNet outperformed the other models for accuracy, precision, and recall. However, it did show the smallest improvement of all the models. The Adapted ResNet50 again outperformed the ResNet50. The Simple CNN received higher accuracy but lower recall than the ResNet models.

I selected the AlexNet model to be my final model, with an accuracy of 0.7775 and a recall of 0.7423 on the testing set. On the validation set, it scored an accuracy of 0.8013 and a recall of 0.7734, a substantial improvement over the original model.

To gain insight as to how the model is functioning, I created saliency maps for three cardiomegaly images. The saliency map highlights regions of pixels that are highly used by the last convolutional layer to determine the output.



The saliency map reveals that the model particularly utilizes the border between the heart and the lungs to make its determinations, particularly the border on the patient's left. This shows that the model is functioning using features associated with cardiomegaly. The enlargement of the heart changes the angles of these borders, which is then detected through the convolutional layers.

Conclusions and Future Directions

Using chest X-rays, I built an AlexNet model that detects cardiomegaly with an accuracy of 77.75%, precision of 79.85%, and recall of 74.23%. While this model is successful, there is room for further improvement, specifically as it would be deployed in the medical field for diagnosis. A further means of improvement would be training on a larger amount of the NIH X-ray database while implementing data augmentation for the limited 2776 cardiomegaly samples. Additionally, I used the image size used by AlexNet and ResNet of 224 x 224 pixels, but the original images are 1024 x 1024 pixels. The processing time of that amount of data did not fit within the scope of this project, but setting the image size to 320, 512, and 1024 pixels should be explored for future improvement. Additionally, as the Adapted ResNet50 outperformed the original, adapting the AlexNet with more fully connected layers may prove beneficial, with the caveat that unlike the AlexNet architecture, the only trainable layers in the ResNets were the fully connected layers.

As this project confirms that medical imaging can be used as data for machine learning, future projects could explore using X-rays to detect other conditions, including TB and pneumonia. Additionally, since cardiomegaly is in practice diagnosed by the ratio of the heart to the thorax, a project should explore determining the cardiothoracic ratio. This would require more data and specifically bounding boxes around the heart for both patients with and without cardiomegaly.