

# Predicting Pneumonia Using Chest X-Rays

Andrew Yip  
March 2020

# Pneumonia

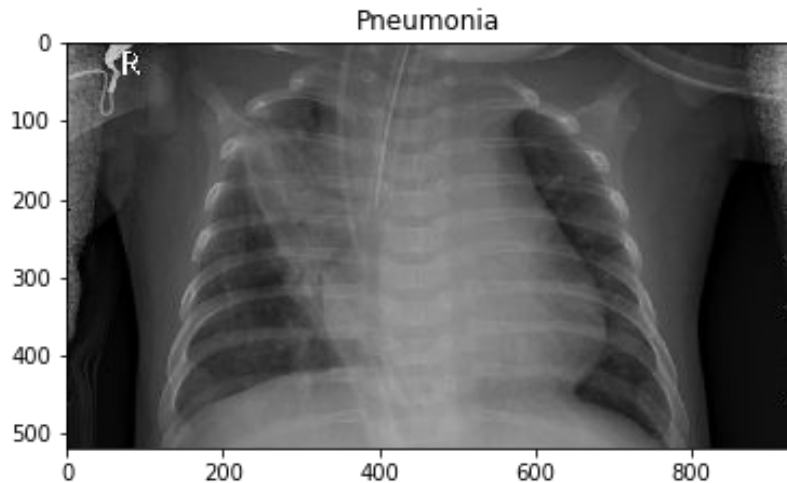
- Affects millions
- Most common cause for hospital admission
- Faster detection, earlier treatment

## Data from Kaggle

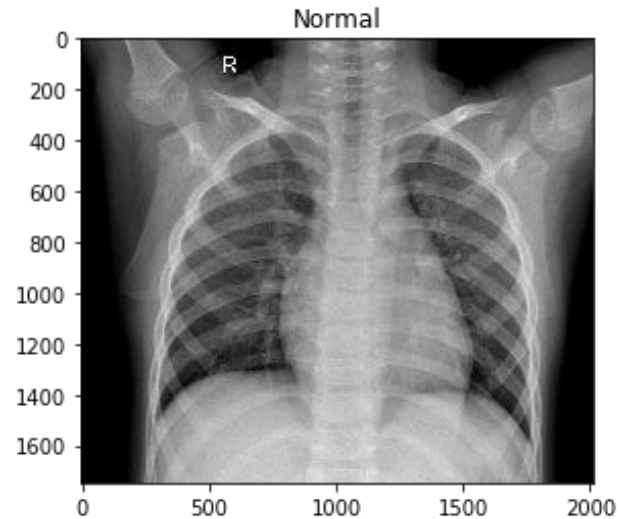
<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

# Chest X-Rays

Pneumonia

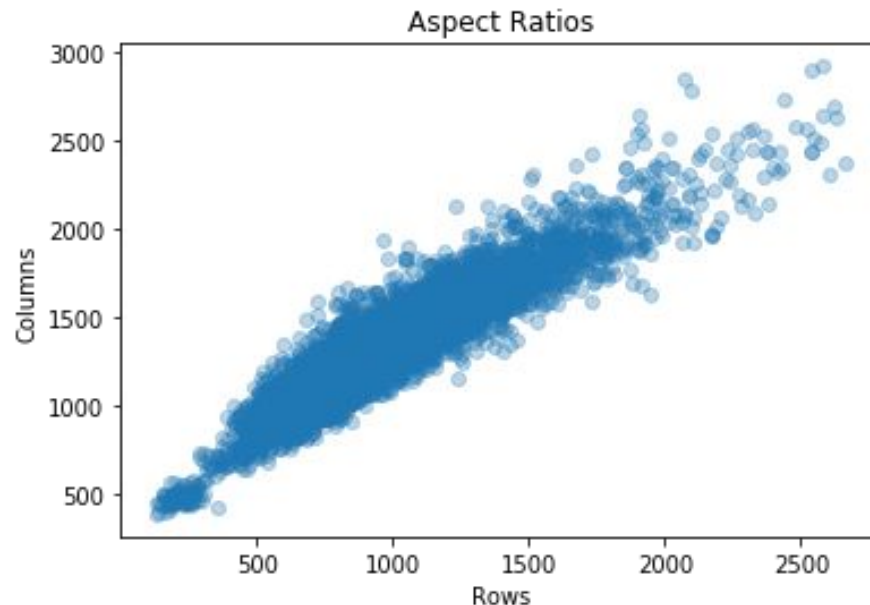


Normal



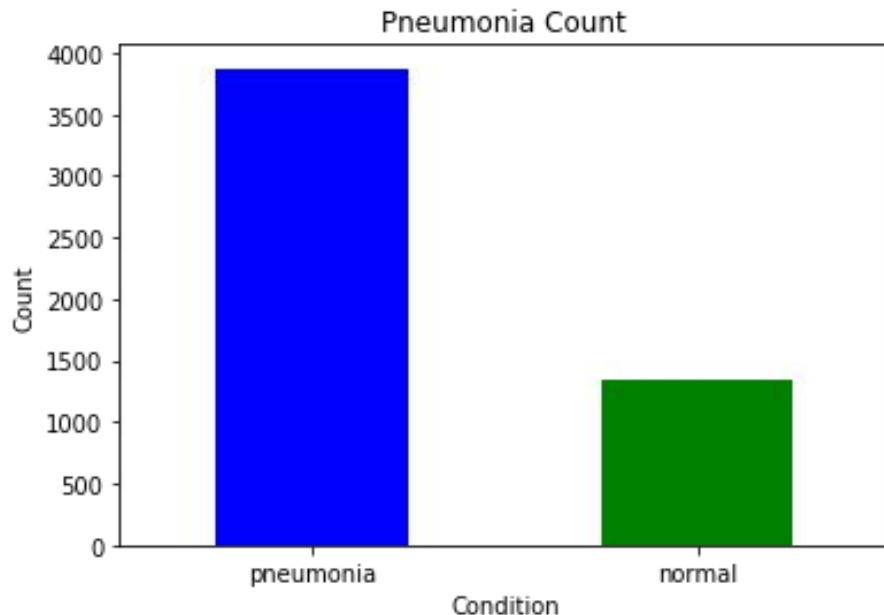
# Image Size

- Images vary in size
- Suggests linear relationship between rows and columns
- Average aspect ratio = 1.44
- Use this ratio to resize images for deep learning model



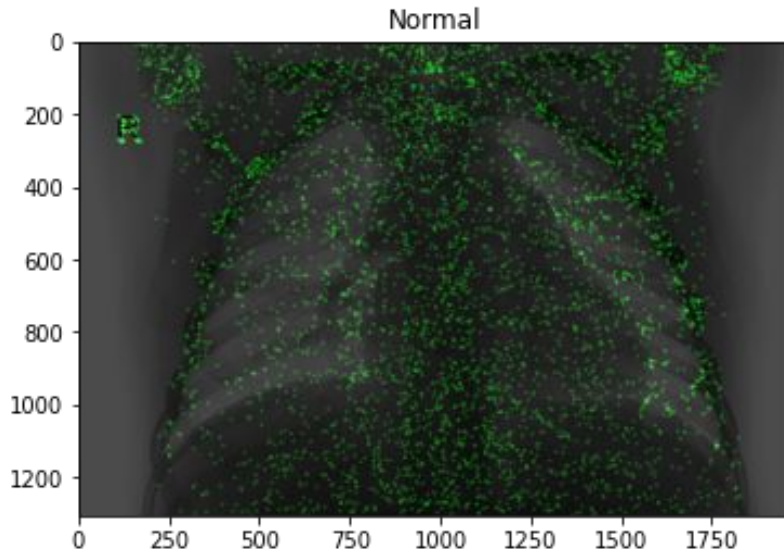
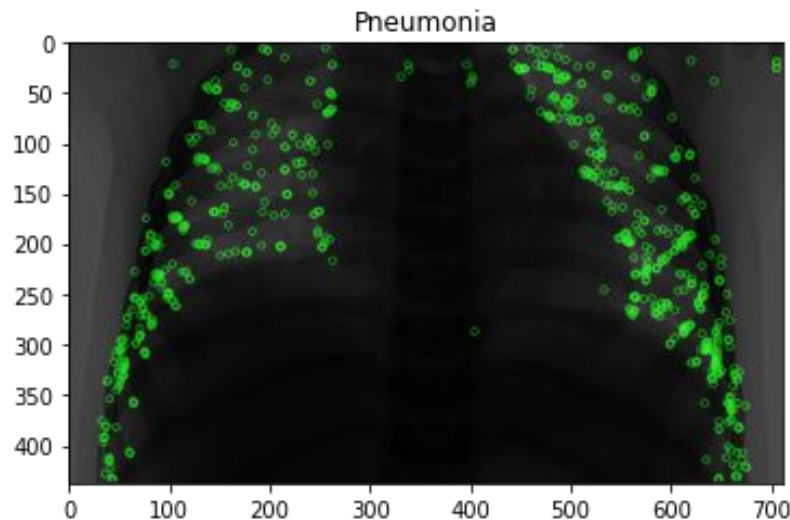
# Category Counts

- More pneumonia images than normal images
- Imbalanced data set
- Use F score, precision, recall to evaluate model



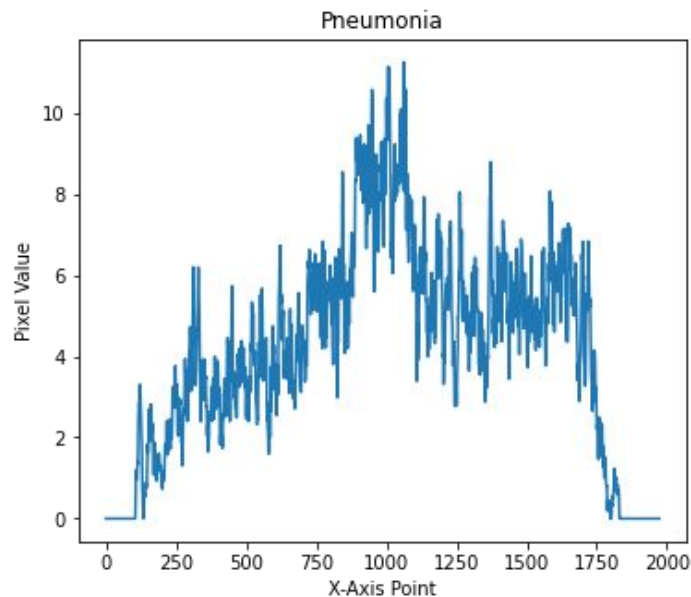
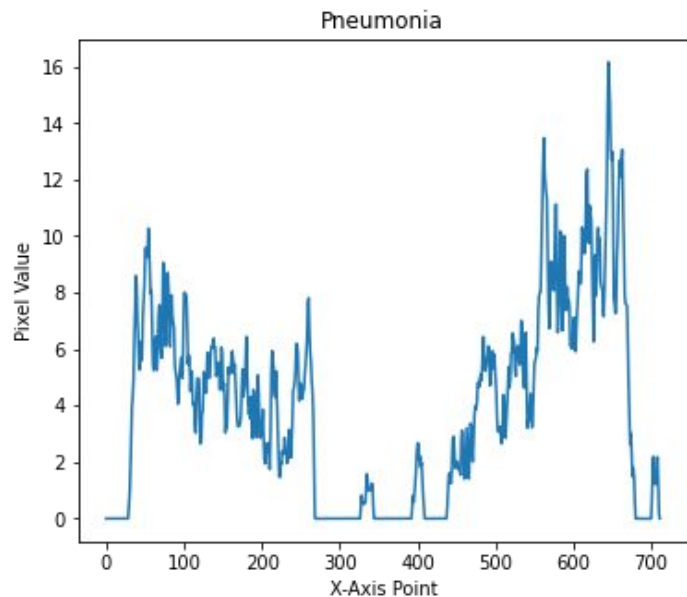
# OpenCV for Image Processing

- How do we tell which image is the one with pneumonia?
- Use FAST algorithm in OpenCV for feature detection.



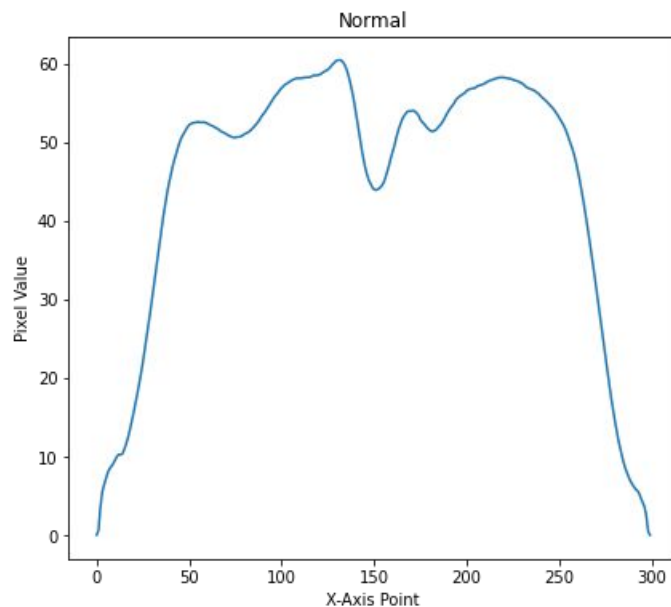
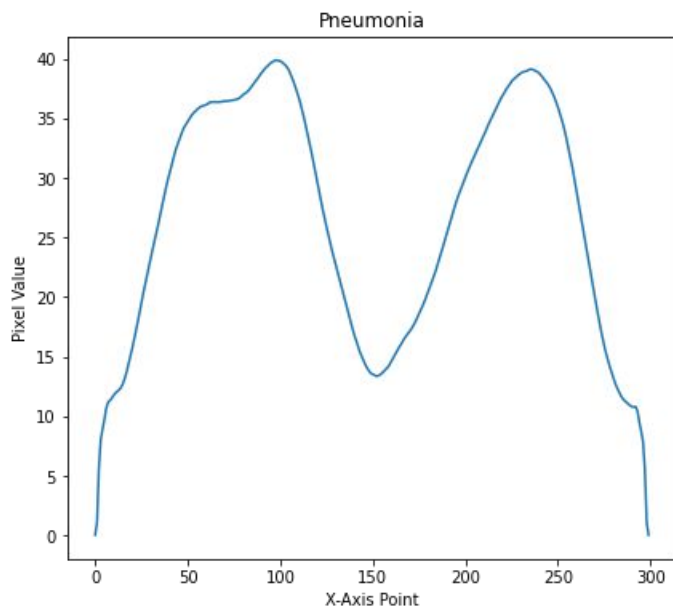
# OpenCV for Image Processing

- Lack of features around lungs
- Dip in average pixel intensity



# OpenCV for Image Processing

- Plot the averages of the mean pixel values across the x-axis





# Building a Model

Use Keras to implement deep learning model with convolution layers.

- After some experimentation with various model architectures, we settle on one that produces desirable results

## TRAINING

Loss: 0.25 Acc: 0.92 F1 Score: 0.94 Recall: 0.89 Precision: 1.00

## VALIDATION

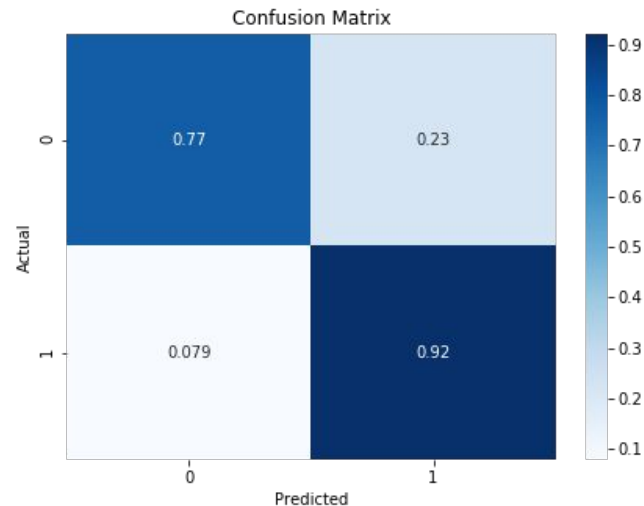
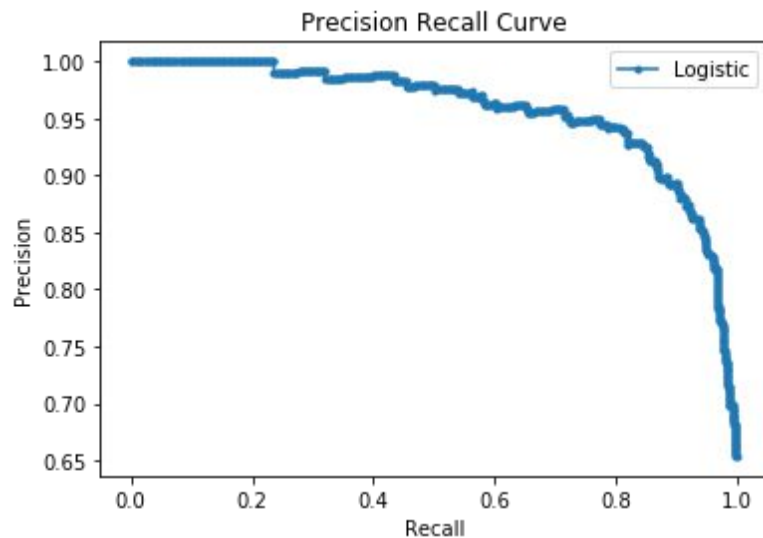
Loss: 0.26 Acc: 0.92 F1 Score: 0.94 Recall: 0.90 Precision: 0.99

## TEST

Loss: 0.37 Acc: 0.87 F1 Score: 0.89 Recall: 0.92 Precision: 0.87

# Model Evaluation

- Validation and Test metrics produces similar results, suggesting we are not in danger of overfitting



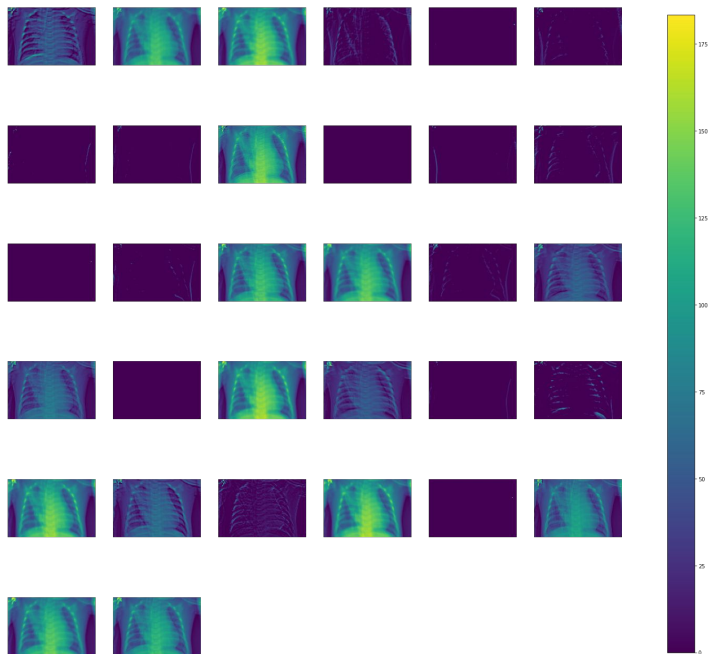
# Interpreting Our Model

Now that we can predict pneumonia, we should seek to understand how our model is making its predictions.

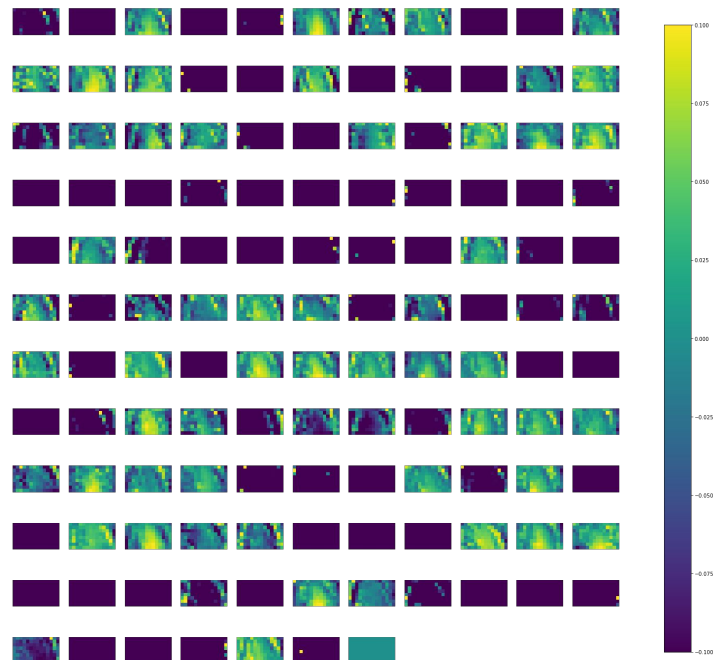
- Visualize our filters and activations for our model's convolution layers
- Look at a heatmap of what areas/features our model sees as more important

# Visualizing Convolution Layers

conv2d\_47

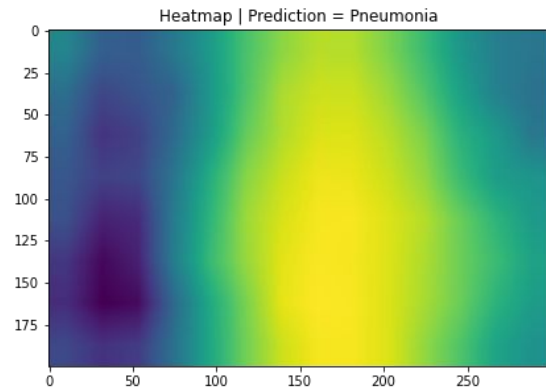
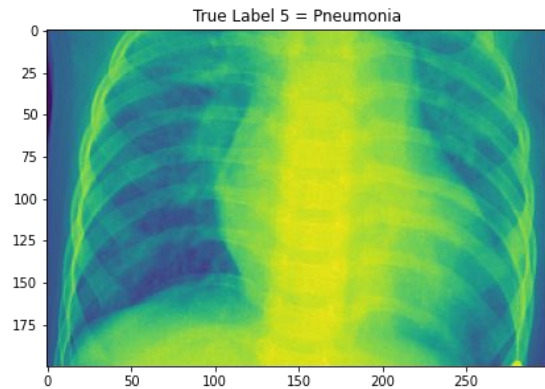
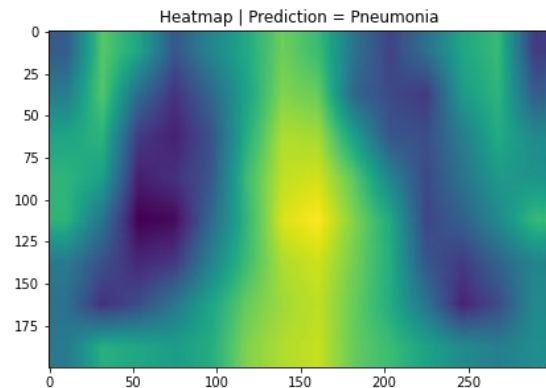
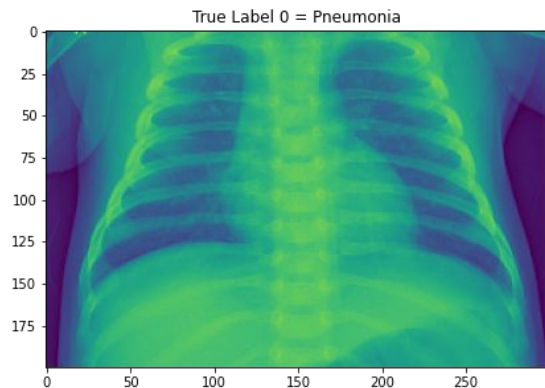


conv2d\_50



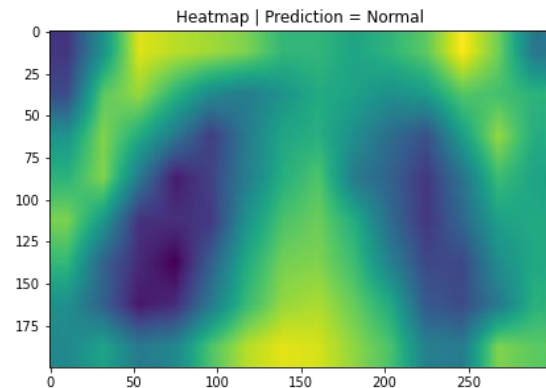
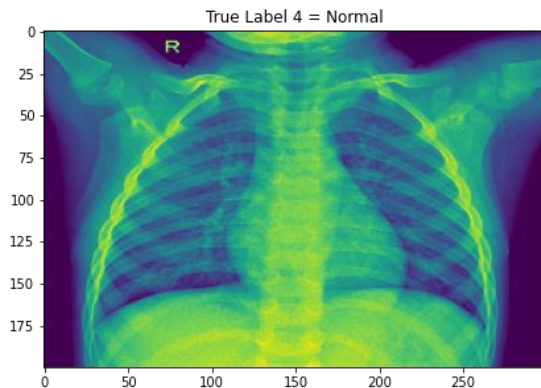
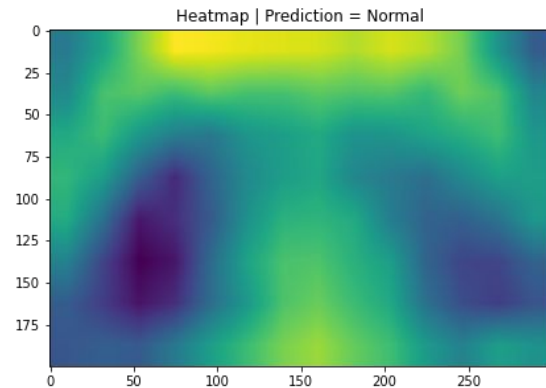
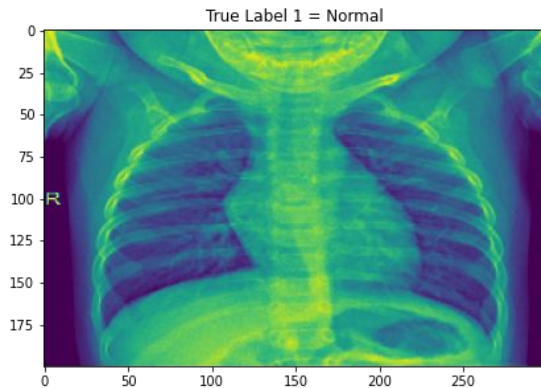
# Grad-CAM Heatmap

Large high intensity area at the center of the image generally results in 'Pneumonia' prediction.



# Grad-CAM Heatmap

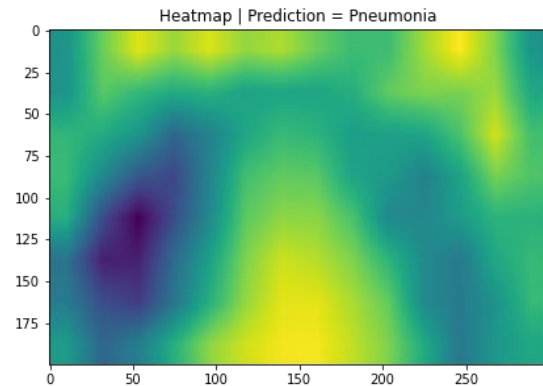
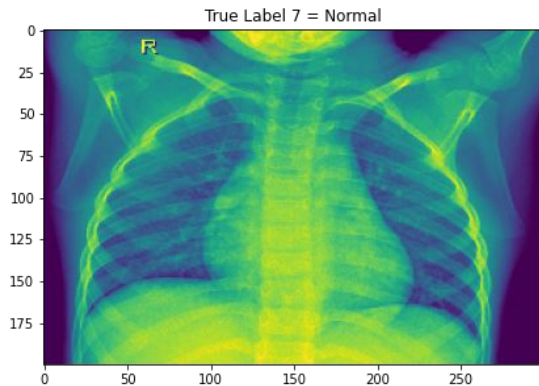
Lack of an obvious high intensity area at the center of the image around the lungs generally results in a 'Normal' prediction.



# Grad-CAM Heatmap

Misclassified as  
'Pneumonia' when it's true  
label is 'Normal'.

This is likely due to the  
detection of an area of high  
intensity at the center of the  
image as seen in the  
heatmap.



# Conclusion

- Have a model that can predict pneumonia from chest x-rays
- Future iterations should keep recall high to avoid type II errors
- Use heatmaps to understand why the model makes certain predictions

# Next Steps

- Further explore more complex architectures to improve results
- Deploy model as a web app