

# Using a Neural Network to Distinguish Between a Healthy Patient and one with Pneumonia Proof of Concept

Freddy Abrahamson

July 1, 2022

# Outline

- Business Problem
- Data Understanding
- Explaining the Model
- Model Results and the metrics used



# Business Problem

- In 2018 there were about 1.5 million people diagnosed with pneumonia in U.S emergency rooms.
- Around 40,000 people died of pneumonia in the U.S in that same year.
- In 2020 pneumonia related deaths jumped to nearly 48,000, with the emergence of COVID-19 contributing to those numbers.

These facts underlie the need for a cost effective, non-invasive diagnostic tool that could correctly identify an x-ray image as one of someone with pneumonia. This could be used to prioritize which patients the doctor should see first, thereby helping save lives, and protect from more severe damage caused by the disease. An additional benefit is that the personnel needed to get these results, does not need to have a medical background.

Data based on the following websites:

<https://www.cdc.gov/dotw/pneumonia/index.html>

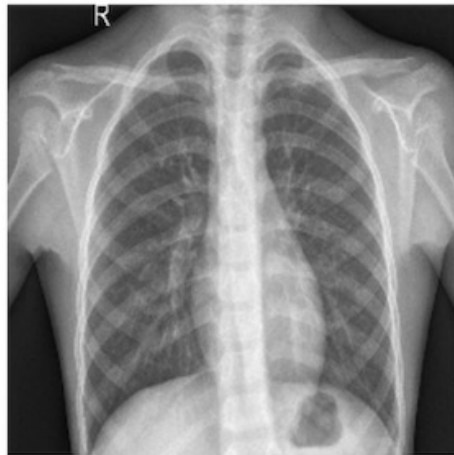
<https://www.thoracic.org/patients/patient-resources/resources/top-pneumonia-facts.pdf>

<https://wonder.cdc.gov/controller/datarequest/D76;jsessionid=F87D0D0AA50B13677E435B758B44>

# Data Understanding

- The data comes from Kaggle.com.
- There are a total of 5856 images.
- This includes 1583 'normal' images, and 4273 'pneumonia' images.
- The ratio of 'pneumonia' images to 'normal' images is about 2.7 : 1.

normal image

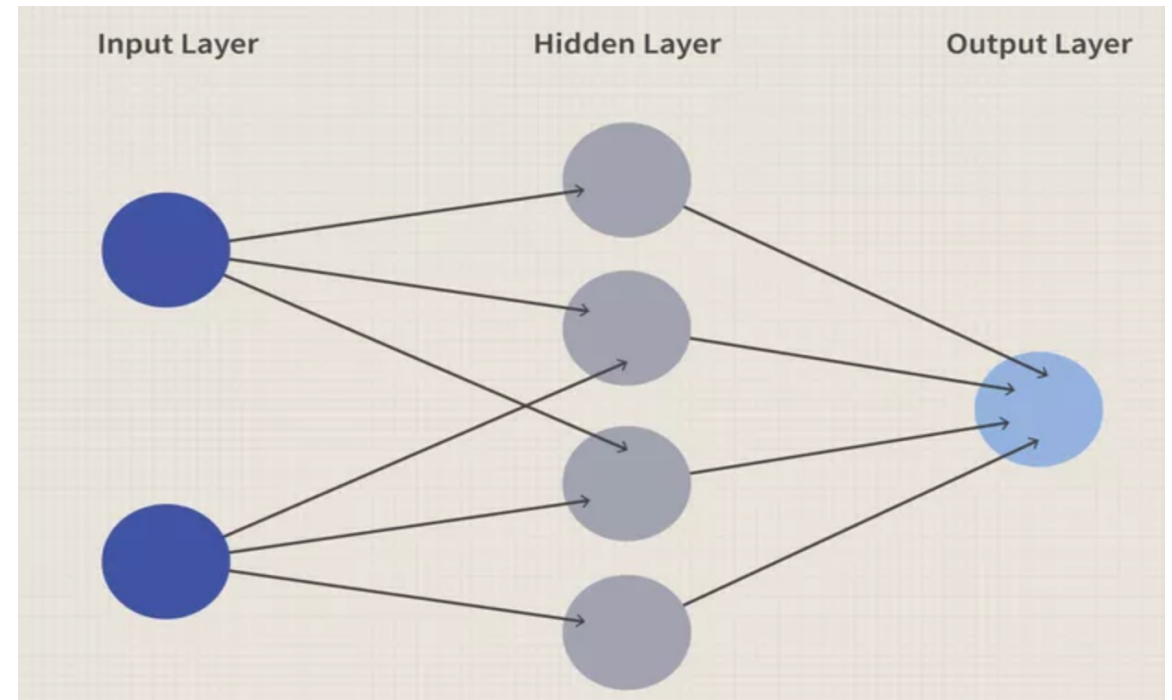


pneumonia image



# The Model: What are Neural Networks ?

- The machine learning model we used is called a neural network.
- Neural networks were inspired by the human brain.
- Just as the human brain is made up of neurons that form an interconnected network and send electrical signals to each other to help humans process information, a neural network is made up of 'artificial neurons' called nodes that similarly share information with each other to help the network process the data.
- The layers are made up of columns of nodes, and the data moves from left to right.
- The networks are very scalable. The number of layers, as well as the number of nodes per layer can be modified.



The circles are the 'artificial neurons' or nodes. The input layer receives the data. The output layer returns a result, and the hidden layer(s) are any layers in between the input and output layers.

# Model Results

- Given that our primary goal for this model was to successfully identify all the patients who had pneumonia, the metric we used was recall.
- The model performed very well, identifying 410 out of a total of 428 patients with pneumonia, which gives us a recall score of nearly 96%.
- In addition to that, the model performed very well with regards to accuracy.
- Accuracy is calculated by dividing the number of correctly identified images , by the total number of images.
- With 559 out of 586 total images, correctly identified, we have an accuracy score of over 95%.
- To recap: The model was able to successfully identify nearly 96% of patients with pneumonia, and was able to correctly identify over 95% of all images.

Confusion Matrix

True label	pneumonia	normal	
	410	18	
pneumonia			
normal	9	149	
Predicted label			
		pneumonia	normal

**Any Questions?**

**Thank You**