Image Pneumonia Classification

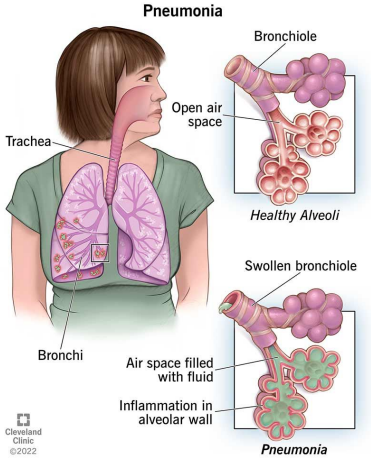
DS4400 Project Proposal

Dave Budhram, Sarah Costa, and Brandon Onyejekwe

What is the problem?

Pneumonia:

- A type of respiratory infection that affects the lungs. It can be caused by bacteria, viruses, or fungi, and it can occur in people of all ages. Pneumonia is a serious condition that can be life-threatening, especially in older adults, young children, and people with weakened immune systems.
- Can we find a way to detect pneumonia given patient x-rays?

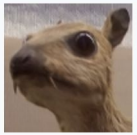
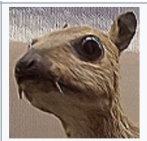


Why is it interesting?

- Instead of using numbers and strings for feature data, we are relying on images instead. This requires less collection of specific features that could be relevant to predicting the outcome
- Tools using machine learning to analyze patient data are becoming helpful tools for medical professionals to use

What are the approaches you plan to explore?

- Utilizing different activation functions
- Testing different numbers of hidden layers and neurons in each layer
- Dimension reduction (PCA) before feeding into a model
- Sharpening and/or otherwise manipulate the x-ray images using convolution matrices to see how that affects the model

Operation	Kernel ω	Image result g(x,y)
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	

Approaches and Limitations

- Manipulating (e.g. sharpening) the image can make some of the important features more pronounced than others, making it easier to classify
  - However it might take a while to run code than sharpens every image in the dataset
- Will split between bacterial and virus vs dataset with both types. Does the combined model detect Pneumonia just as well as the split models?
- Dimension reduction can help the model focus on the main components of the image, but it may take away some of the information that is important to classify the image
- The interpretability of the image themselves since don't have particular features to rely on

Timeline and Division of Work

- April 3rd: finish setting up the model
- April 4th - April 12th: test image manipulation effects, different activation functions, and multilayers

Division of Work

- We plan on setting up the model together.
- Dave: Testing Tanh activation function
- Sarah: Testing Sigmoid
- Brandon: Testing ReLU
- After finding the best activation function we'll work on testing using different layers together
- If time permits: add images after manipulation to the dataset and repeat the above process