

Data Appendix

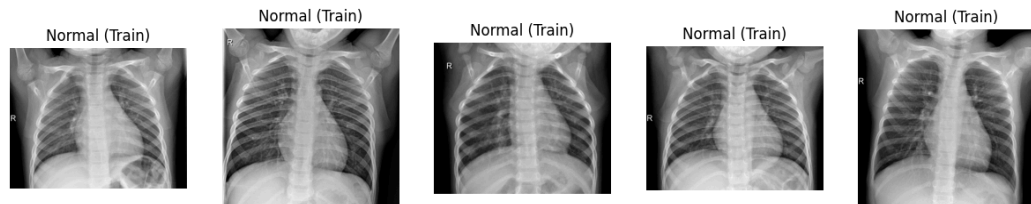
For this project, we analyzed data from the [Labeled Optical Coherence Tomography \(OCT\) and Chest X-Ray Images for Classification](#) Dataset, specifically focusing on the ChestXRy2017.zip file. This appendix serves to provide more information about this dataset

There were no missing observations for any of the variables, as this is an image based dataset, so the variable descriptions will not mention this for the sake of simplicity.

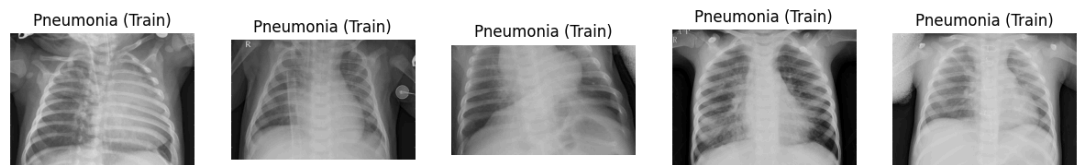
Details of Each Variable:

Images were split into Test/Train dataset, using a 20/80 split.

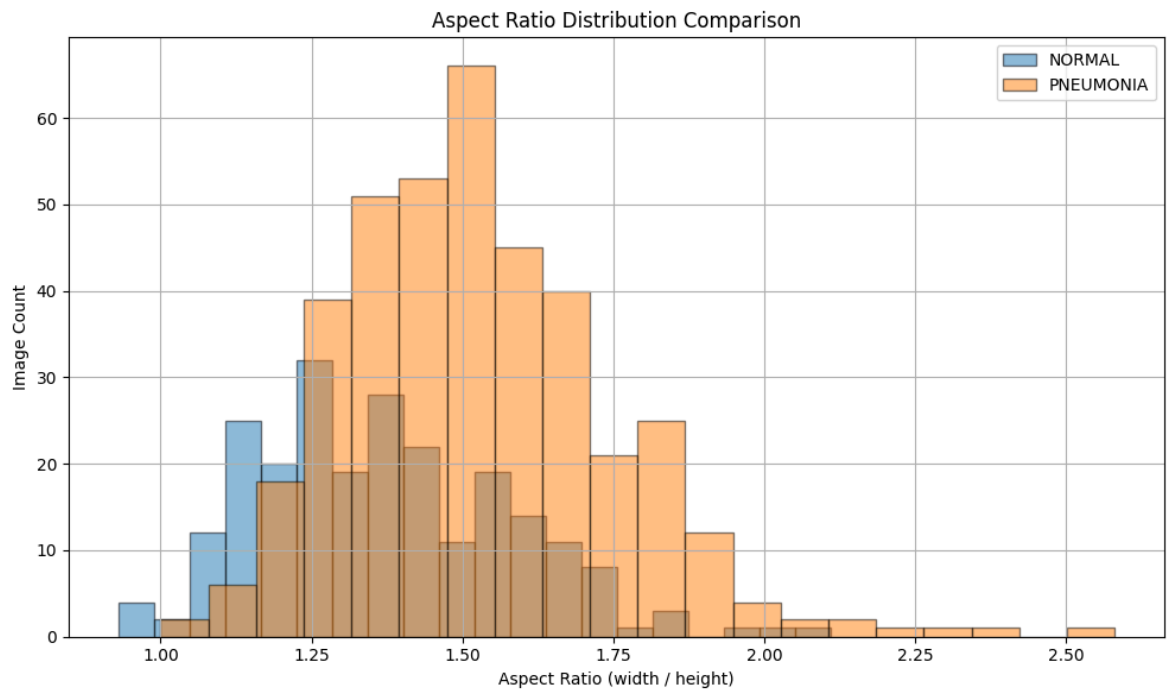
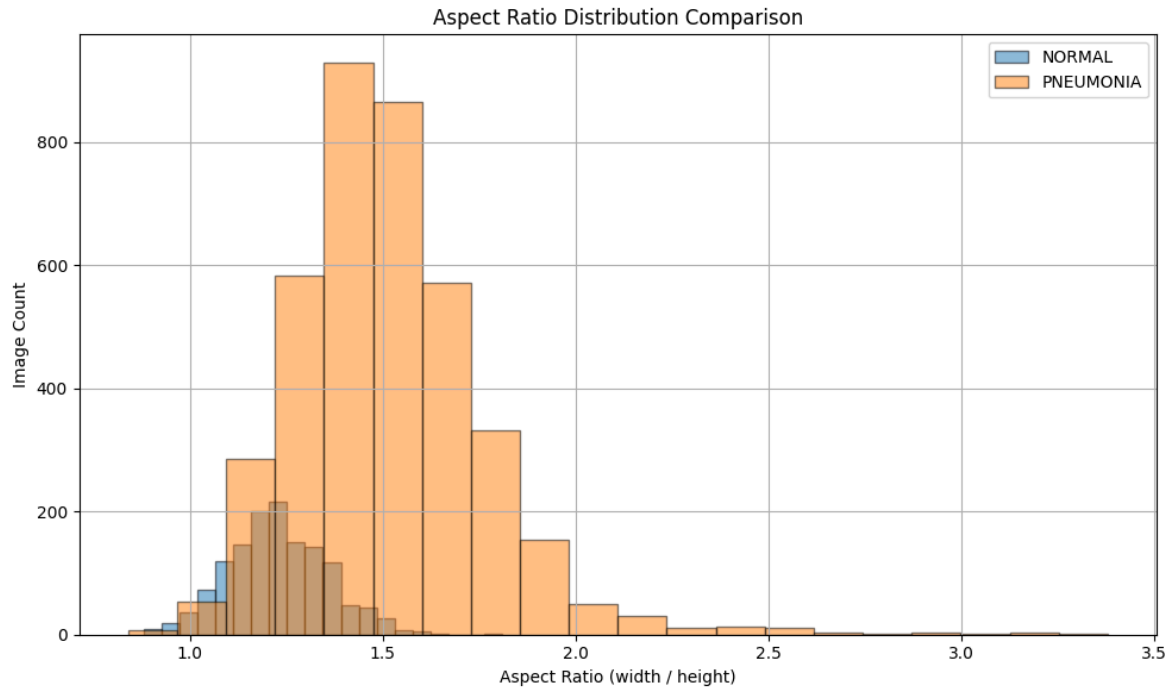
- Normal
 - Definition: X-ray Images of healthy patients, i.e. exhibiting no pneumonia
 - Data type: Image
 - Processing steps: Resized to 384x384



- Representative images of Normal patient X-rays from the Train dataset
- Pneumonia
 - Definition: X-ray images of patients with pneumonia
 - Data Type: Image
 - Processing steps: Resized to 384x384



- Representative images of Pneumonia patient X-rays from the Train data set



-
- Aspect Ratio Distribution in NORMAL vs PNEUMONIA X-ray images in the Train and Test Dataset