**Initial Ideas: Sept 19, 2024**

**Data Wrangling / Cleaning:**

* Likely won’t have too much data wrangling or cleaning to do, as the data provided to us are images
* Since part of the model will be telling the user if the image quality is good or not, we don’t want to exclude poor images from data
  + Might eventually exclude them from our data once we build the model, but this would be if we are specifically building a model that deals with high-quality images (or, at least, good-enough quality)

**EDAV:**

* Look at distribution of good images, images with spinal cord and fluid, and other labels between the old dataset and the new dataset
  + Want to see how many images have only fluid, only spinal cord, spinal cord and fluid, or neither

**Other Ideas:**

* Want to investigate why the model performed poorly on the new dataset vs the old dataset
  + One idea is to use saliency maps: see what regions of the images are being used by the model to make predictions
    - Compare between the old set and the new set
* Can mess around with adjusting the images: stuff like zooming in, histogram normalization, shearing, AutoAugmentation, salt+pepper noise

**Questions to ask:**

* What exactly is the image quality label telling us? Is it whether or not the spinal cord / fluid is visible, if the image quality itself is acceptable, or something else?

**Goal for week of 23rd September:**

1. Get data access and do initial pass of what the data looks like (brief EDAV, can later divide up the work for the report)
2. Investigate what models they have already been using and see how they work / how well they work
3. Maybe try some of the easier modifications to the models, like adjusting the images or incorporating a siamese architecture

Sequential?

Weed out bad images, and then work only with the good ones?

Overlap with summer work

New labels? Do we consider them or not?

Nerve roots?