Paisley Lucier
Presentation 1 - GitHub documentation

I. **What did you do?**

A) I created visuals for representing Luna's ideas for using different color representations (slide 7).

B) I reviewed last quarter's train / validation / test split in the second sheet: ML model comparison. I examined how we might move forward with (1) identifying the types of patterns that are crucial to be in the training data via clustering and (2) stratifying the training data accordingly to see if there is an improvement in recall. Note: during the presentation this was amended to potentially consider using attention to identify which cases to stratify instead of clustering.

II. **How does it help the project?**

These tasks were part of the initial process of identifying where we want to spend our efforts this quarter and potential first executable ideas to improve the MIL process from last quarter, and additionally, they helped me get a baseline understanding for moving forward.

III. **Issues faced (if any)**

There was lots of information to go through given the many quarters of prior work, and I had trouble figuring out what to focus on–ie. identifying a starting point on what to improve. Additionally, the information online that is relevant to this project is quite complex, so I admittedly had a lot of trouble reading through and understanding research online.

IV. **Attempts to resolve issues (if any)**

I leaned on what others in my group were already working on, referred back to notes on what we discussed in mentor meetings / lectures leading up to the presentation, and I also reviewed my STAT 362 notes on CNNs to better prepare for understanding the work to come.

V. **Issues resolved (if any)**

No issues were resolved this week.

VI. **Next steps**

Next steps for the project include:

(1) Writing code to implement the updated MIL model

(2) Writing code to experiment with alternate color representations in MIL

(3) Choosing a method to categorize patterns (eg. using attention to identify which features really matter, or clustering), and then implementing an experiment to test the impact of associated stratification

VII. **References**

Slides (Presentation 1 - STAT 390)

Last year's results (ML model comparison)

StratifiedGroupKFold documentation
Maximilian Ilse, Jakub Tomczak, Max Welling: Attention-based Deep Multiple Instance Learning (understanding attention weights to use to identify important patterns)
ChatGPT (starting point with slide 7 visuals, ideas on effectively stratifying)

VIII. Links to the presentation & code

Project 1, Presentation 1 slides link: Presentation 1 - STAT 390
No code was written this week.