Paisley Lucier
Presentation 4 - GitHub documentation

I. **What did you do?**

A) I ran the model (Jack's most recent model, which includes weight decay) with an elastic net penalty on each layer of attention. The results were not good, hence their exclusion from Presentation 4. However, using penalties on attention weights in combination with weight decay could be something we further explore (eg. changing the L1 / L2 penalty ratios).

B) I created 4 additional train / val / test splits, where each case appeared in the test cases for exactly one of these splits. I added these .npz files to MIL_trainer_17Nov_Paisley and trained the model 4 additional times using the 4 new splits. I then collected the results and data visualizations for these results.

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

A) was an idea discussed with the teaching team as a way to counteract the negative effect of hierarchical attention, so it was important to try it out and see if the model performance improved. B) was crucial to determine whether the high performance seen in the recent adjusted model in split 1 translated into other train / val / test splits. Evidently, the first split's performance exceeded the other splits, and thus more tuning / adjustments are needed.

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

Firstly, it took a few tries to develop the splits correctly: with each case appearing in one test set, and with stratification of the train, validation, and test sets. Then, once I started running the iterations, I had repeated issues with overwriting (and unfortunately wasted much computational power). Given our current naming convention, it was also quite time-consuming and difficult to match runs to splits to results as I worked through the above issues.

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

I unfortunately ended up running the training multiple times. However, after the second time, I realized that checkpoints were not being overwritten, so I could dig through the folder to access those for each split, which is how I ultimately got the results on the slides.

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

I was able to use the checkpoints in training to visualize and compare results for each of the 5 splits.

VI. **Next steps**

Next steps for the project include:

(1) Changing the naming convention for runs

(2) Tuning the hyperparameters in the training configuration

(3) Further investigation into attention disparities / adding penalties to attention weights

VII. References

ChatGPT (conceptual understanding help, code assistance)

VIII. Links to the presentation & code

Project 1, Presentation 4 slide link: Presentation 4 - STAT 390

Project 1, Presentation 4 code link: https://github.com/arvindkrishna87/STAT390_CMIL_Fall2025/tree/main/Code/MIL/MIL_trainer_17Nov_Paisley