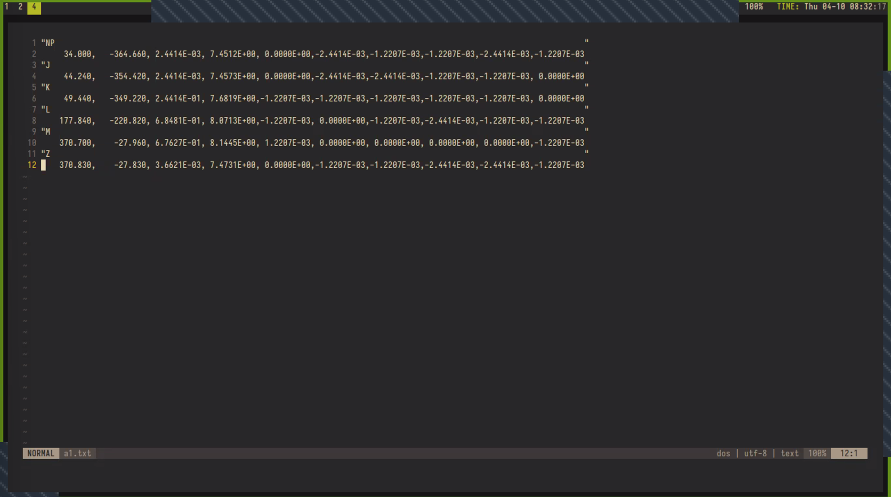
CS USDA-Auburn Meeting Minutes

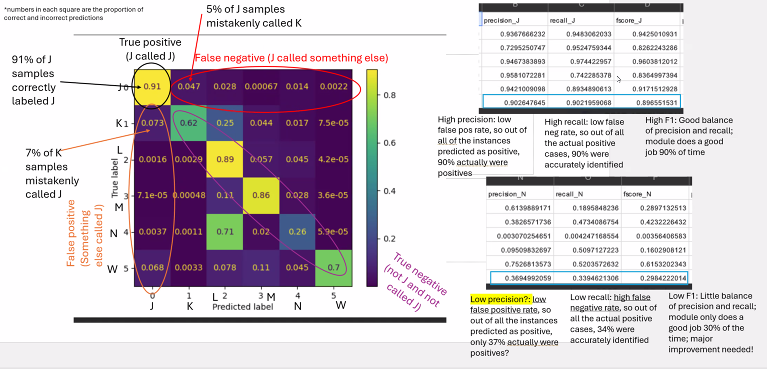
4/10/25



On output format:

* WINDAQ generates this .txt file
* We would be fine outputting just the label and time (first two columns) without voltage
* Dr. Reif: why not just retain voltage as well?
* Dr. Backus resolves we should not–we will focus on outputting label and time

Moving on to ML discussion with Dr. Cooper



* Dr. Cooper:
  + Looks like 26% of the time, N is labeled correctly. 71% of the time, N is labeled as L. So there is a false negative. How do these numbers relate to precision and recall? Why are they both low?
  + Milo: this is not taking the precision and recall from each fold and creating a final result, but taking a simple mean precision and recall without considering the number of data points.
    - There are also weights applied to files (each fold has the same number of files, but different numbers of probes). How do we set these weights appropriately?
    - Devanshi’s interp.
      * Milo: We've done a simple average. How do we want to weigh this? Each of the folds has the same number of files. These are different from the number of probes. Some files have been upweighted as some probes are short.
      * 3 measures: Weight by file, probe, data under each fold
  + Almost has general summary points completed, will send shortly
* What are folds?
  + Zach: We divide data into training and testing sets, so we can test the model on novel data that it hasn’t seen before. Say we happened to have a really easy or really hard test set. We don’t want that, because then our accuracy results would be unrealistic. So in each fold, a different section of data will be used for the test set.
  + We produce a confusion matrix for each fold. The final one is a combination of all these matrices in some fashion to provide an overall summary.
* Reminder from Dr. Reif: concepts and terms that wouldn’t need to be defined to a CS audience will need to be defined in our paper, since it’ll reach biologists, etc.
* Prof. Hope: Confusion matrices are summed and then normalized ?
  + Unnormalized confusion matrices can also be provided, but might be less interpretable
  + Dr Reif: panelized data may be useful.
* On figures:
  + Dr. Cooper:
    - Table 1 will be the summary of the literature
    - Table 2 - I like the one you guys produced, overall accuracy and f1 score for the model
    - Then you had a summary table
    - I combined all of those together that has all of the output averages and put that with each model
    - We’re allowed to have 8 tables in the paper
  + Zach: We wouldn't need to report each individual fold
  + Dr. Cooper: We could have 1 figure showing our 5 overall confusion matrices
  + Dr. Cooper was also thinking about doing a multipanel figure. 2 or 3 of True, predicted, labeled parallel figures.
* Dr. Cooper will be reaching out with methods questions, but for now is trying to make sense of the figures
* Was post rectification used?
  + Milo: Random Forest used post (confirm?) and UNet used pre, TCN used pre
  + Dr. Backus: Did we smooth the waveforms before using them?
  + General concern is that we are not using rectified data
  + Milo addresses this: We tried pre and post rect for all models, but found that the pre rect was slightly better. Providing an unfiltered set provides more information than a filtered set. Neural nets are powerful enough to learn and apply any filters they’d need.
  + Dr. Cooper: we need to be very clear on which rectifier is used for what model
  + Zach: it might be worthwhile to talk to the engineering team about these signal concepts–they’d probably be more knowledgeable than us
  + For the random forest model, it could matter whether we use pre- or post-rect. But for the other models, it doesn’t particularly matter because it will adapt. More information is better in these cases.
* Comparison between models
  + “Like comparing apples to oranges” but we said we’d made a way to compare them
    - Zach: Before, RF performance was measured by its success at labeling each second of data (is this true or just example?). Another model might be measured by its success at labeling each hundredth second. So to make them comparable, we can just multiply time windows to set the metrics to the correct scale.
* Segmentation Transformer Model:
  + Just another model we worked on, but was finished very recently so hasn’t been brought up very much
  + UNet is also a transformer. We just don’t use the term “Transformer” very often in our internal documentation. There is a segmentation transformer model *inside* of the UNet.
* Macro F1 Score in Overall table will change
* UNet Report has a paragraph ending in “Additionally, we.” If this was supposed to say something, update it (or delete it if not)!
* Need to adjust confusion matrix code to produce letter labels along the sides
* On describing confusion matrix results:
  + When I’m describing the results from the confusion matrix, can I say that ‘90% of the time it correctly identified J? Or 90% of the samples were labeled correctly for J?’
  + Zach: 90% of the time, J was correctly labeled at J.
* On precision and recall:
  + Both range from 0 to 1. So they can be treated as percentages. 30-40% precision/recall is pretty low.
  + Milo: expected precision/recall is also dependent on the problem. They are also more relative than absolute, as in they reflect a comparison to other models/data, and not a global or flat comparison
* Prof Hope: Might be worthwhile to look at precision/recall curves for each class
* Dr. Cooper is feeling like we’re in much better shape. There are some things to revisit later, but we are ok for now.
* Dr. Backus has a request for the next meeting:
  + She is in the process of manually labeling a very large sharpshooter dataset.
  + Wants to check that we think this is being done in a way that is ideal for the summer interns who will be adapting our work to sharpshooter data?
  + Concern that sharpshooter waveforms are in two groups: very typical, and very diverse/different–so one label could look different between two instances, and there are great variations in waveform frequency. And highly variable numbers and frequencies of probes. Will a model be able to handle this? How will it affect the program?
    - Zach: This might just be a hard problem. There’s nothing Dr. Backus can now do to improve this. The interns will figure something out.
* Dr. Cooper is free to set up a meeting with the ML people to finalize everything!
* We’ll also talk about whether pre- and post-processing has/will be included in our final product.