# Auburn USDA CS Liaison Meeting Agenda

**Date:** 13 February 2025

**Zoom Info:**

* <https://hmc-edu.zoom.us/j/91402147478?pwd=bau13Kp335ceA9YCc3AApWyI5eH6rD.1>
* Meeting ID: 914 0214 7478
* Passcode: CLINIC2025

**Meeting Goals**

* Demonstrate progress towards goals

**Agenda**

* Short logistics discussion regarding site visit [5 mins]
* Project Updates
  + GUI Progress [Mehrezat, Devanshi]
  + WinDaq Comment Importation [Lillian]
  + Data Augmentation and Pre/Post-Processing [Milo]
  + CRFs and Standardized Benchmarking [Zach]

**Assigned Notetaker:** Lillian (Zach backup)

**Minutes**

Site visit

* Clinic admin confirmed travel plans, van reserved
* Prof Breeden recommends Hampton Inn in Selma, meet at Pea Soup restaurant
* Leave monday after classes, all day Tuesday available, citrus plant weds morning
* Traffic bad, will want to plan times carefully

**Project Blurb**: Researchers at the USDA and Auburn University use a technique called electropenetrography (EPG) to better understand the feeding behaviors of arthropods, in particular mosquitoes. Manual analysis of EPG data is time-consuming for our researchers and they seek to automate the process. To help them automatically analyze this EPG data, we created a set of machine learning models they can use through a graphical user interface.

GUI progress - Mehrezat/Devanshi

* Letter labels that move with colored regions
* Began work on a settings panel to adjust colors, grid lines, etc.
* Dr. Reif: how does the information get exported to a data file?
  + For our programs we’ve been using densely labeled files, but we’ll export timestamps of transitions
  + Dr. Backus: would also be good to have individual durations (deltas between interval end times)
  + Dr. Cooper: all the statistical software is based on ending time of the intervals (begin with NP) — would be good to do this too, for consistency
  + The
* Summary statistics is on our todo list

Extracting comments from WINDAQ - Lillian

* Lillian got comment importation from the WINDAQ files working, showed a proof of concept with comments at each time
* Plan is to incorporate this into the GUI, Dr. Cooper liked the format used in the proof of concept with comments overlaid on the waveform.

CRF (conditional random fields) - Zach

* Training is very slow, perhaps not ideal for this use case — may need to rethink
* Made script to generate plots and statistics for model in a standard way
  + Precision, recall, F1, accuracy, confusion matrix
  + Will be helpful for the publication

Unet cleanup - Milo

* Testing preprocessing steps - “data leak splitter” and “simple splitter”
* Testing using voltage only and voltage + clasp
* Data leak splitter + voltage only has best performance

Data augmentation - Milo

* For ML with images, one might rotate images as a data augmentation method
* Want more data, so augmentation lets us artificially increase the amount of data we have
* How can we combine waveforms in a way that makes biological sense?
* Dr. Reif: waveform order should be maintained, could maybe replace an “L” from one recording with another — a kinetogram could be useful for this
* Dr. Cooper: appearance of waveforms being dependent on earlier waveforms mainly happens when the wiring job is poor, appearance of “K” is especially different, fluctuations in the baseline — only want to do augmentation with “good” files
* Dr. Backus: for example, M is a repetitive waveform, pumping and swallowing liquid, frequency and amplitude vary across individual insects based on insect health — sharpshooters might vary 2 to 6 Hz, avg 4
* Dr. Reif: if the researcher introduces a variable like a treatment that might influence waveform characteristics, need to be able to detect that variation
* J to L happens ~2% of time, happens only a few times in our data set, so the model will have very strong bias against and will never predict
  + Could artificially increase weight given to this transition when training so the model is more likely to predict
  + Dr. Cooper: adding treatments may cause these rare events to happen much more frequently, so this would be useful to have
  + Dr. Backus: often, rare behaviors become much more common in pesticide-treated insects
  + Dr. Reif: increasing/decreasing waveform duration, multiple occurrences, should predict ways that waveforms might change as a result of treatments
  + Dr. Backus: R/emf tradeoff, changing input resistance changes amplitude of waveform components based on whether they come from resistance or biopotentials
* stitch 2 Ls together, or stretch one