Time-series Modeling, Analysis, Interface, and Insight from Entomological Electropenetrography

Auburn-USDA

Team: Zachary Traul (TL-S),
Devanshi Guglani,
Milo Knell,
Lillian Vernooy,
Mehrezat Abbas (TL-F)
Advisor: Prof. Gabriel Hope
Liaisons: Dr. Elaine Backus (USDA),
Dr. Anastasia Cooper (Auburn),
Dr. Kathryn Reif (Auburn)

Outline

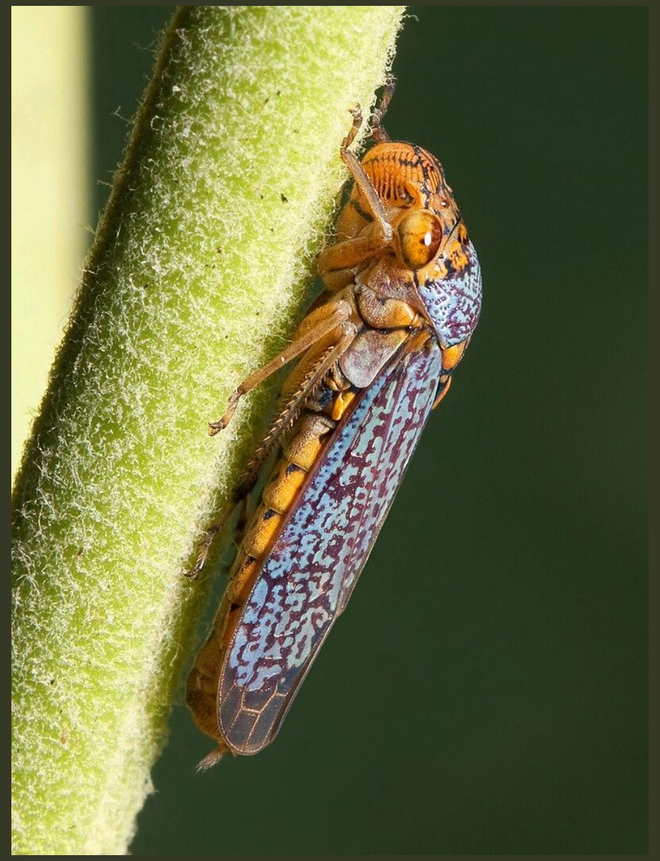
01 Background

02 Project Goals

03 Accomplishments

04 Future Work

Aphid

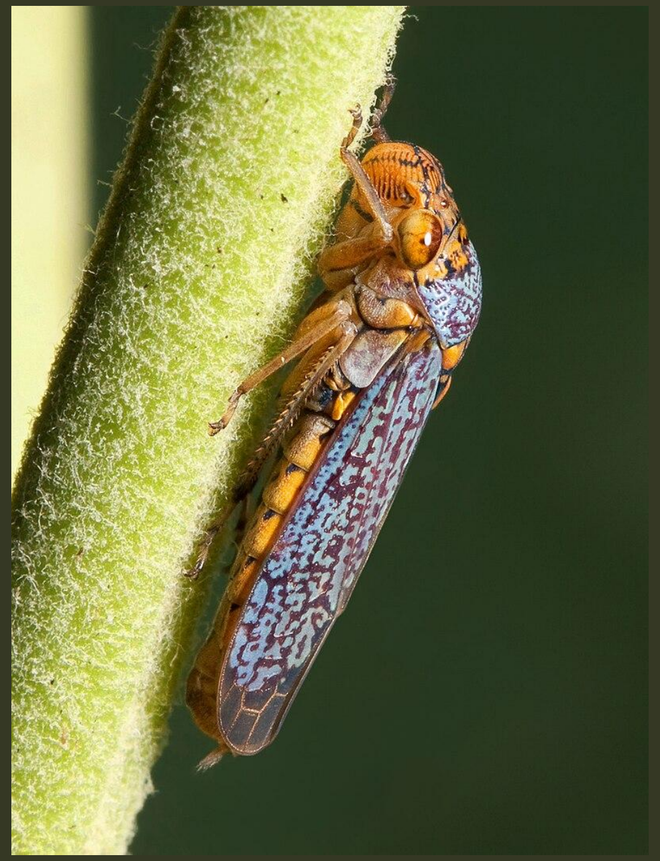


Sharpshooter

Mosquito

Pierce’s Disease caused by Sharpshooters (University of California)

Aphid



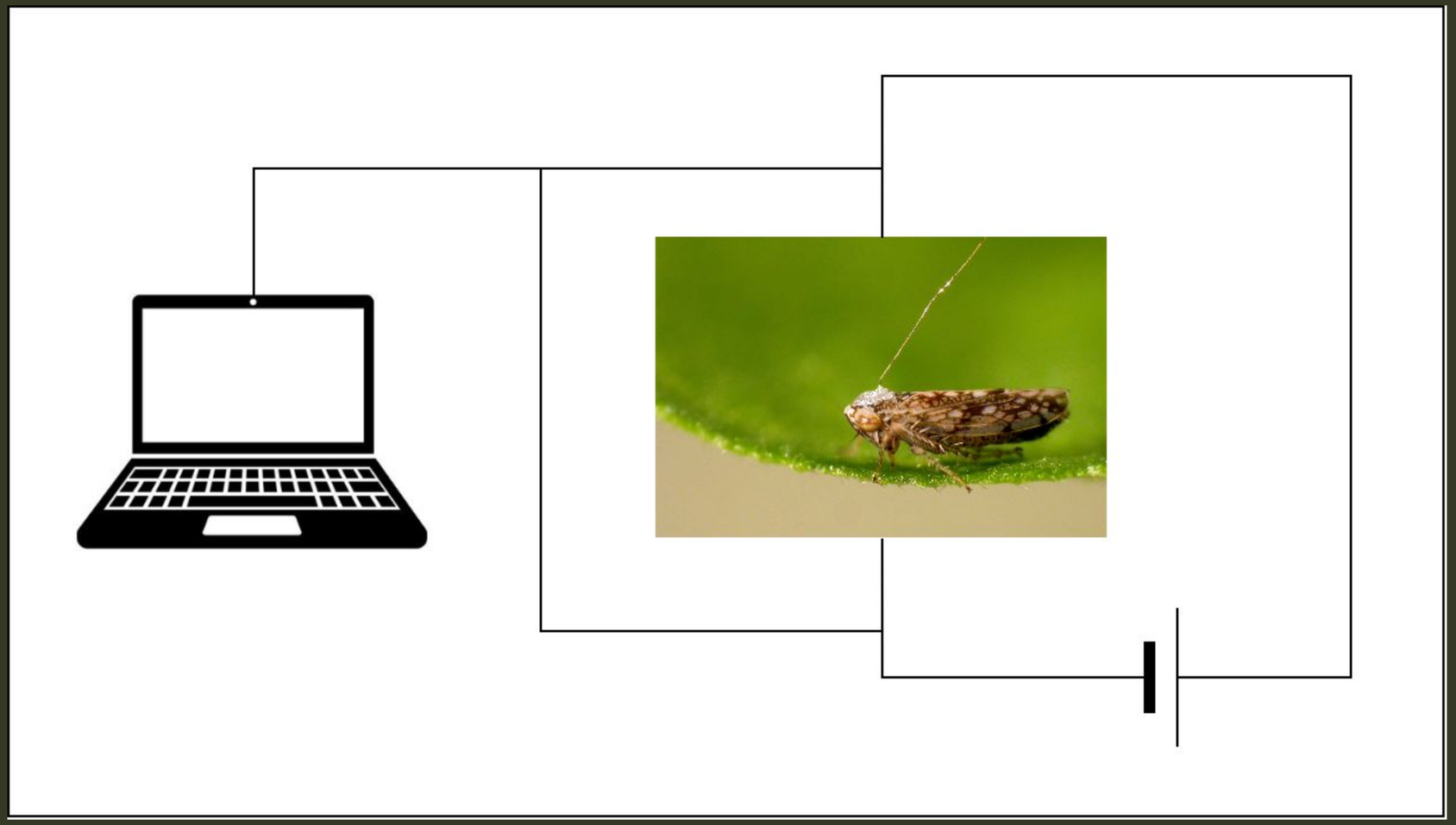
Sharpshooter

Mosquito

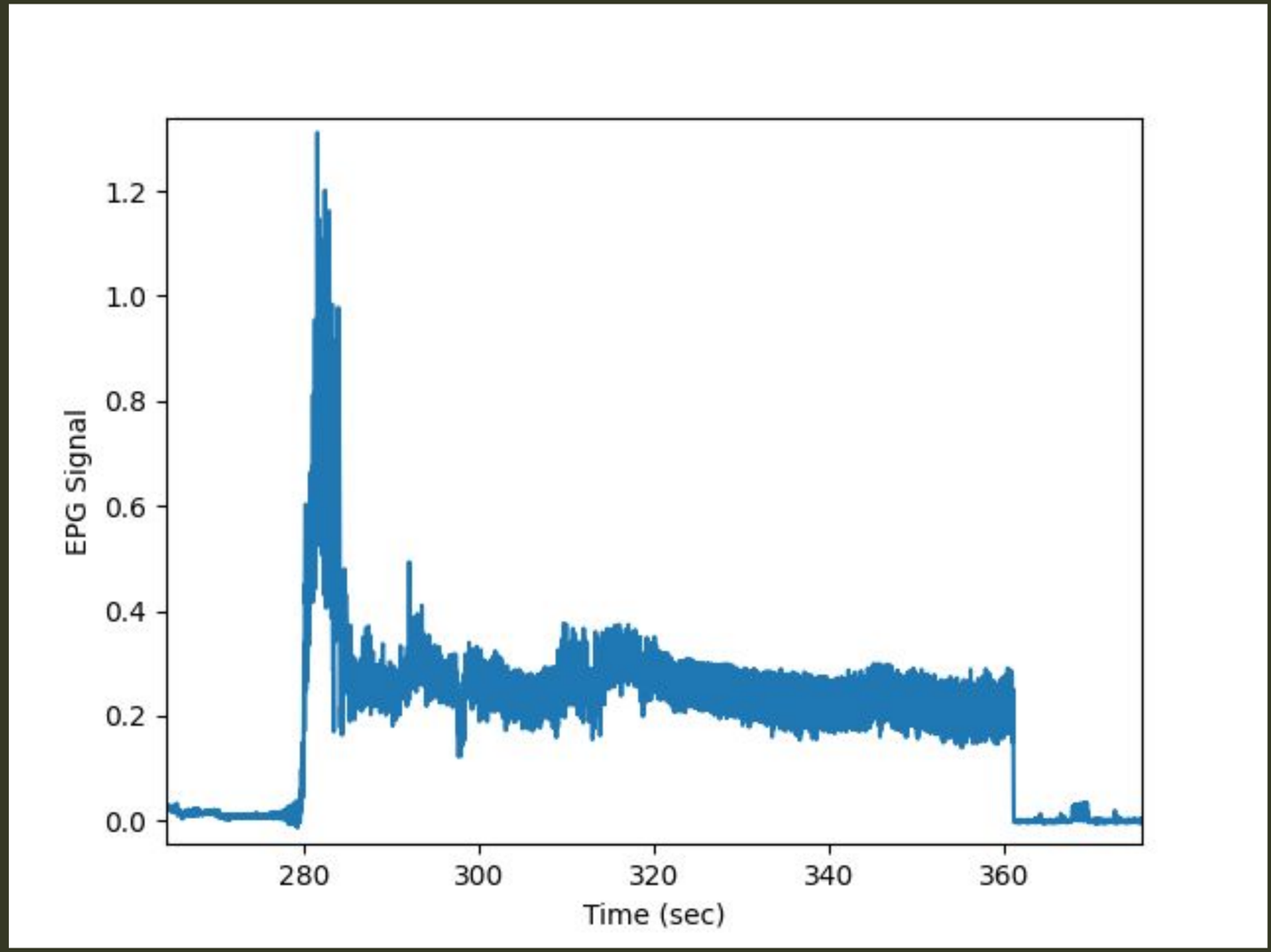
We can’t directly observe what the mouthparts are doing

Electropenetrography (EPG)

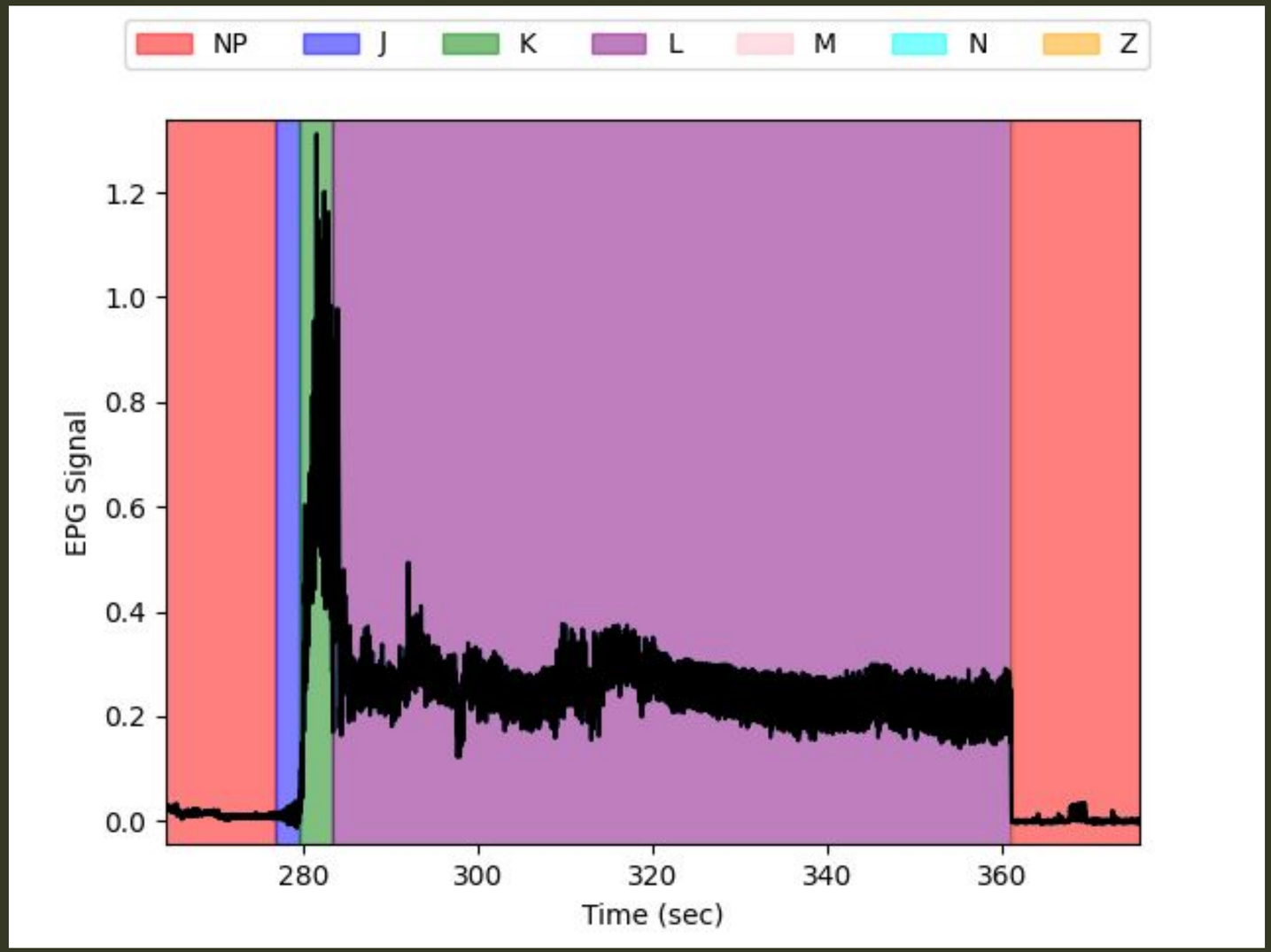
Leafhopper ready for EPG



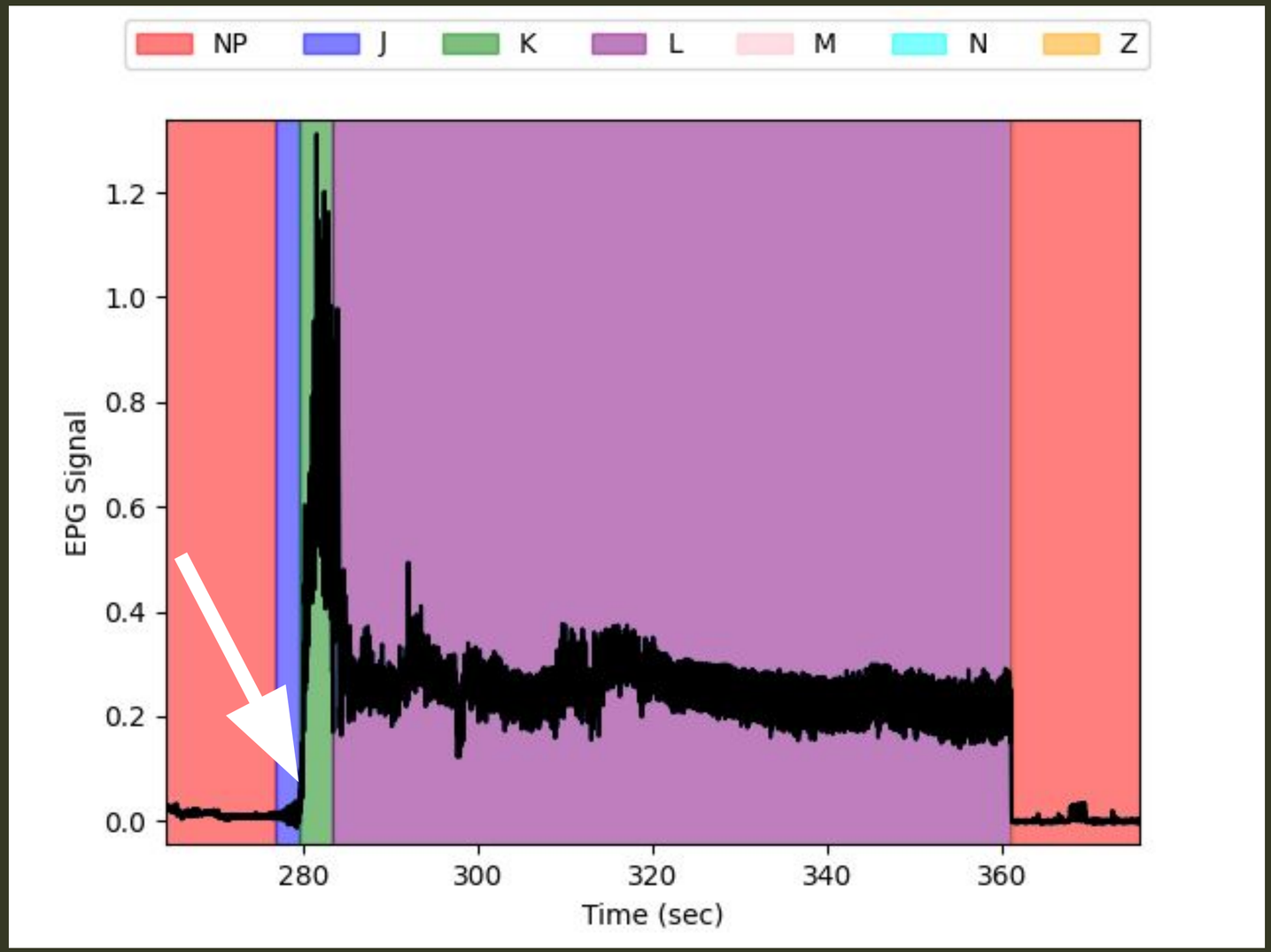
EPG Circuit



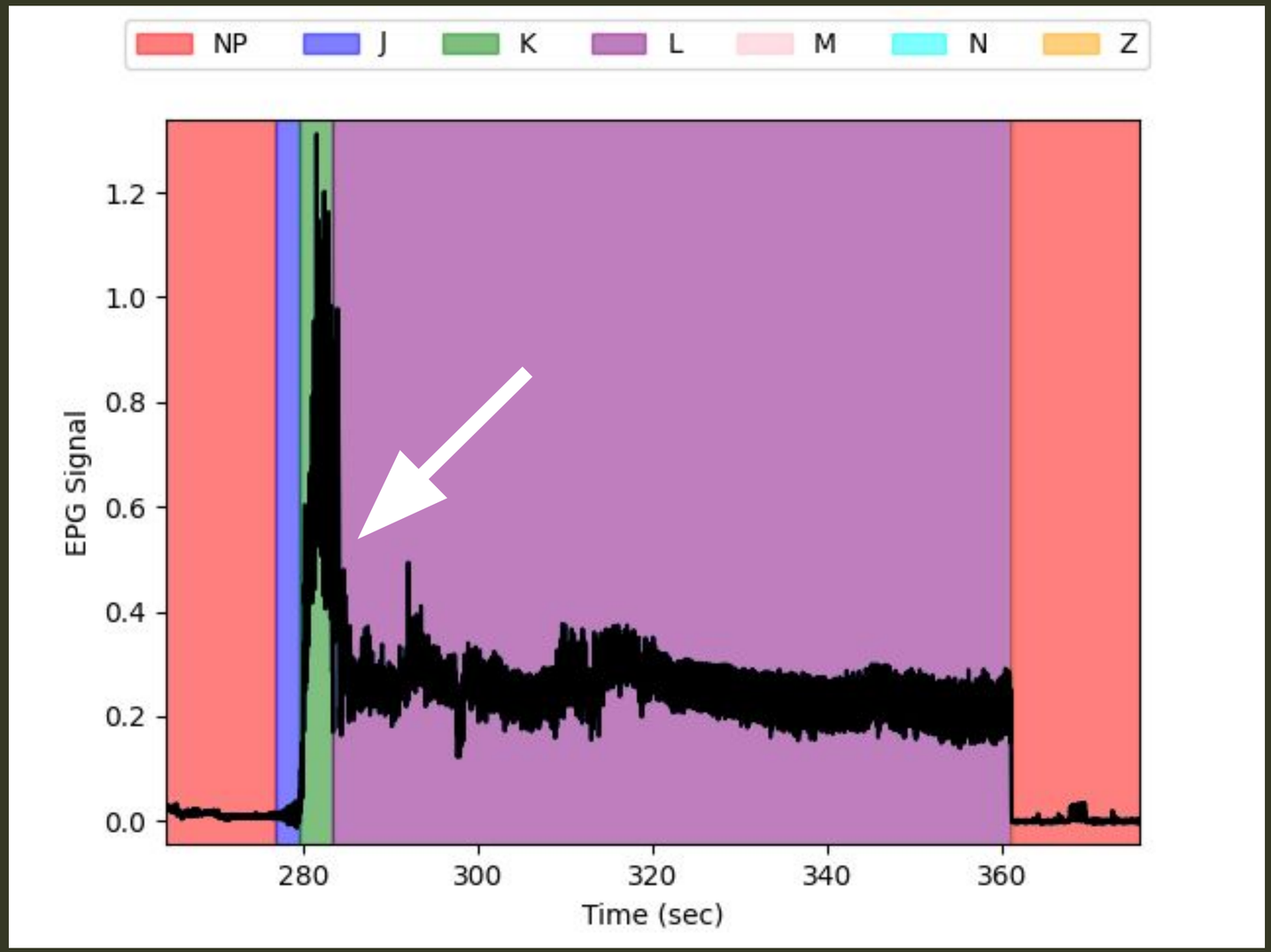
EPG Recording



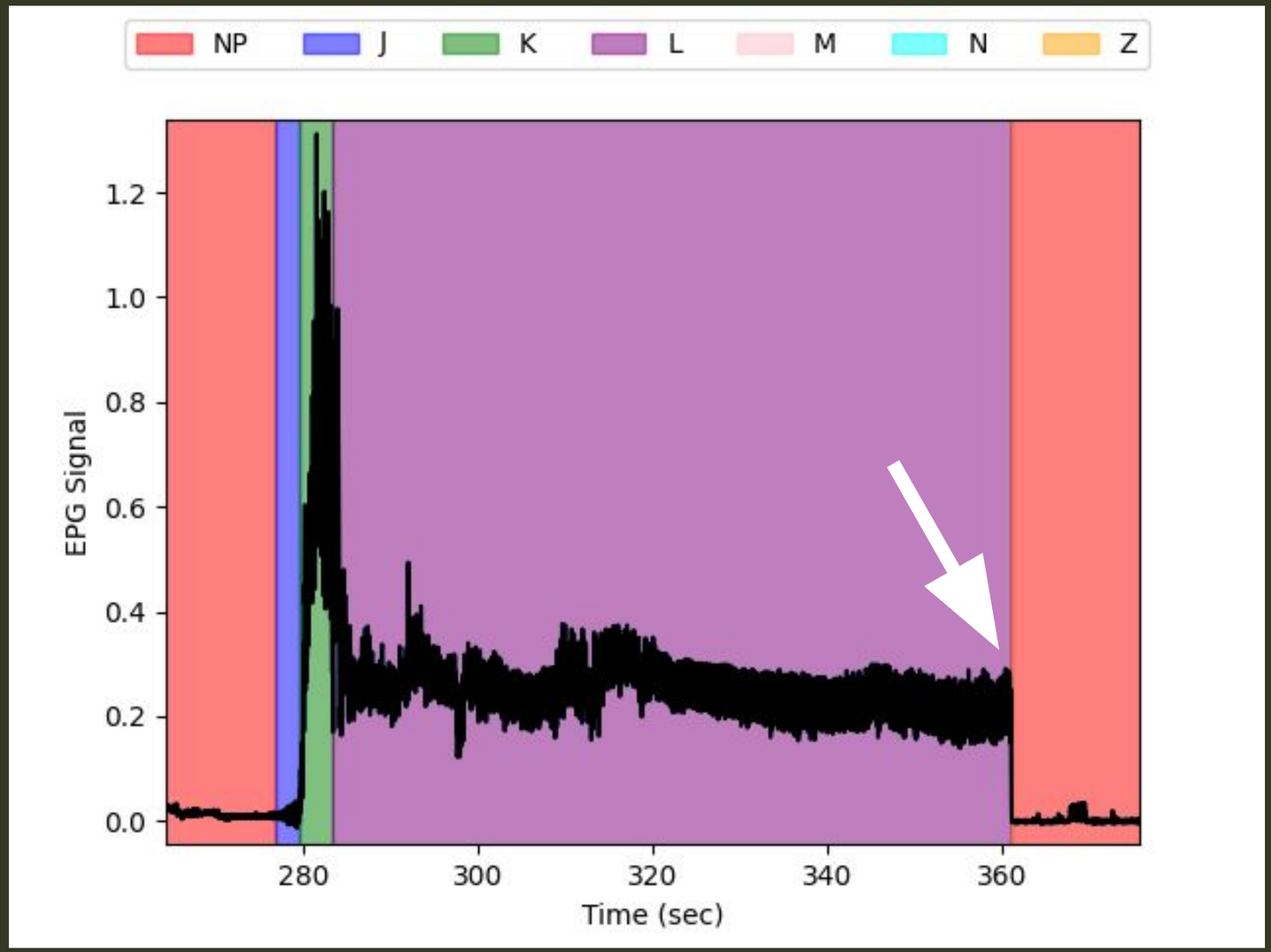
EPG Recording



EPG Recording



EPG Recording



EPG Recording

Our task: Automate EPG labelling and make it accessible

Deliverables

Train predictive ML models for waveform segmentation

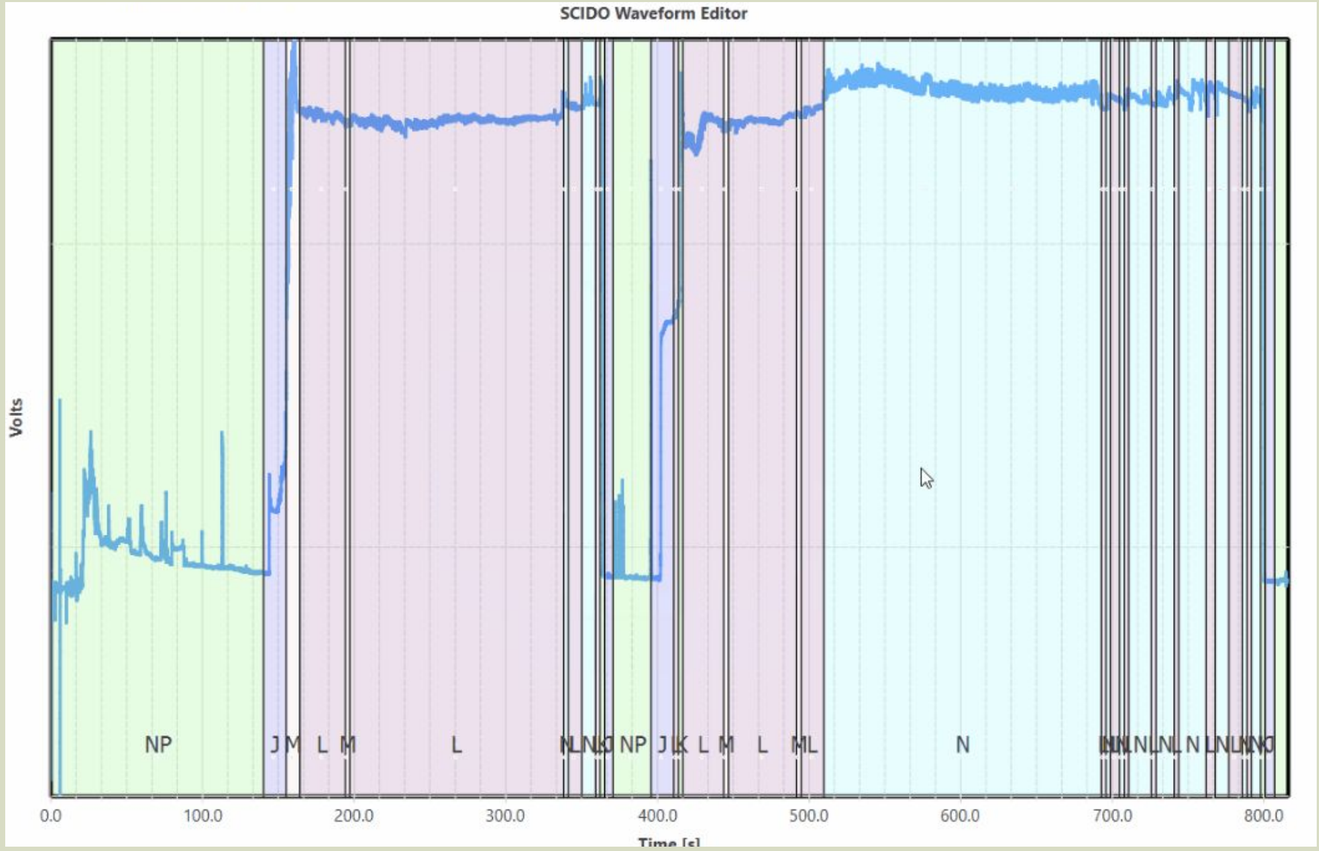
- Explore approaches to automated EPG labeling
- Accurately label EPG recordings
- Integrate seamlessly with GUI

Present it with an accessible user interface

- Simple visualization of data
- User oversight of the automated labeling
- Tools for manual labeling

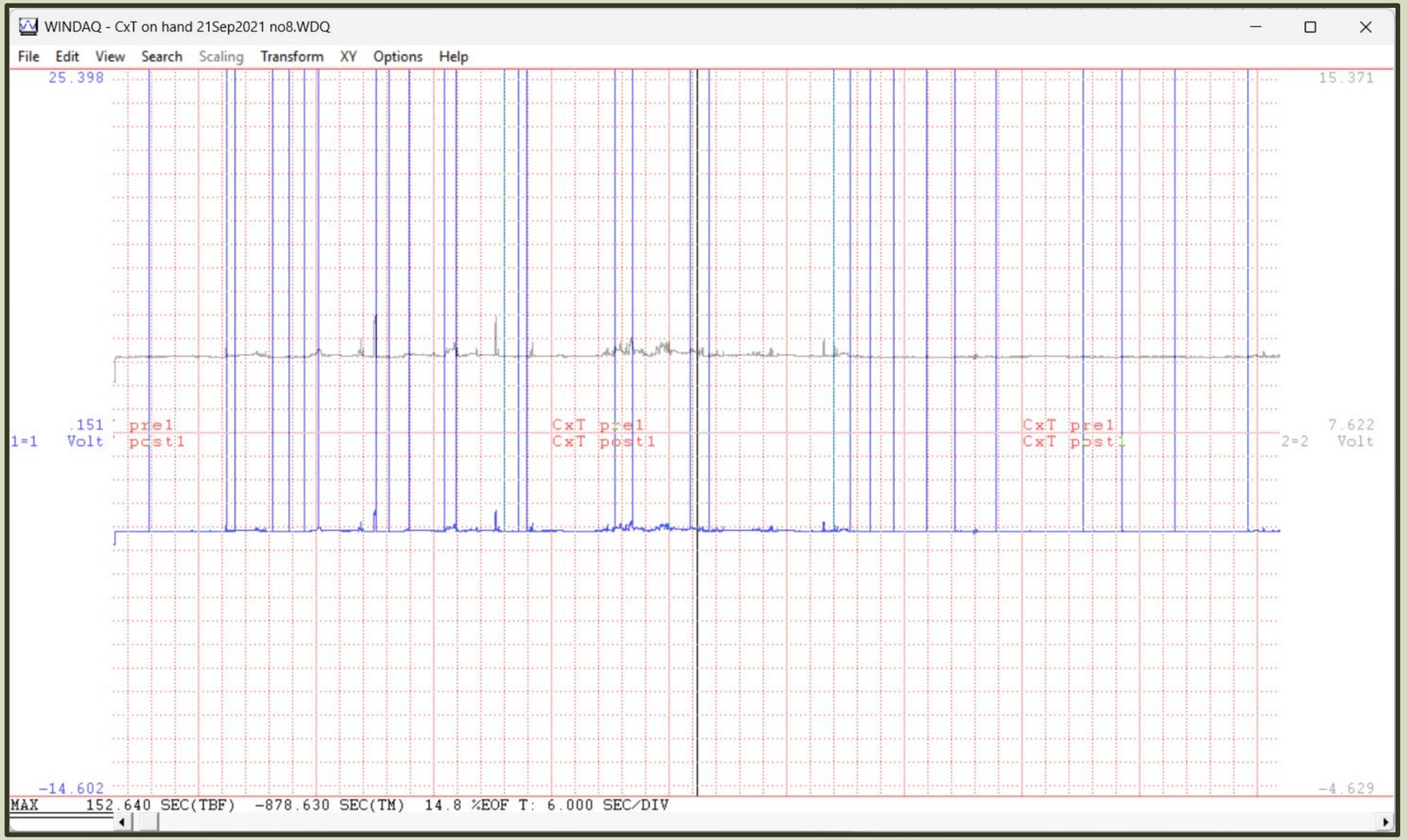
Why ML?

- Labeling is a tedious process for humans
- For automated recognition, we need:
  - Consistency
  - Speed
- Makes it perfect for a ML model!



Why do they need a GUI?

- Windaq is inefficient and cumbersome
- Doesn’t work with ML
- Not extendable
- Can’t plug models in



Workflow

Visualization (data-to-user)

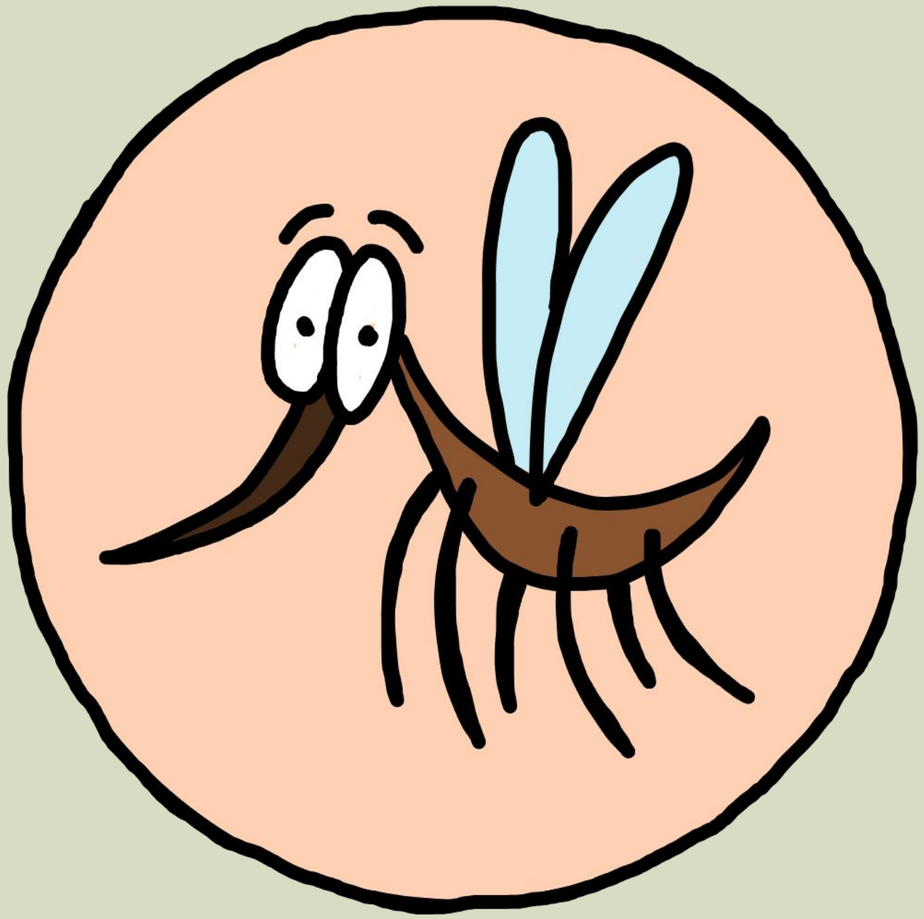
- Labeled EPG data in time series
- Color-coded regions highlighted
- Overall modernized experience compared to Windaq

Characterization (user-to-data)

- Apply the ML model to data
- Adjust, delete, modify labels
- Characterization without alterations to dataset

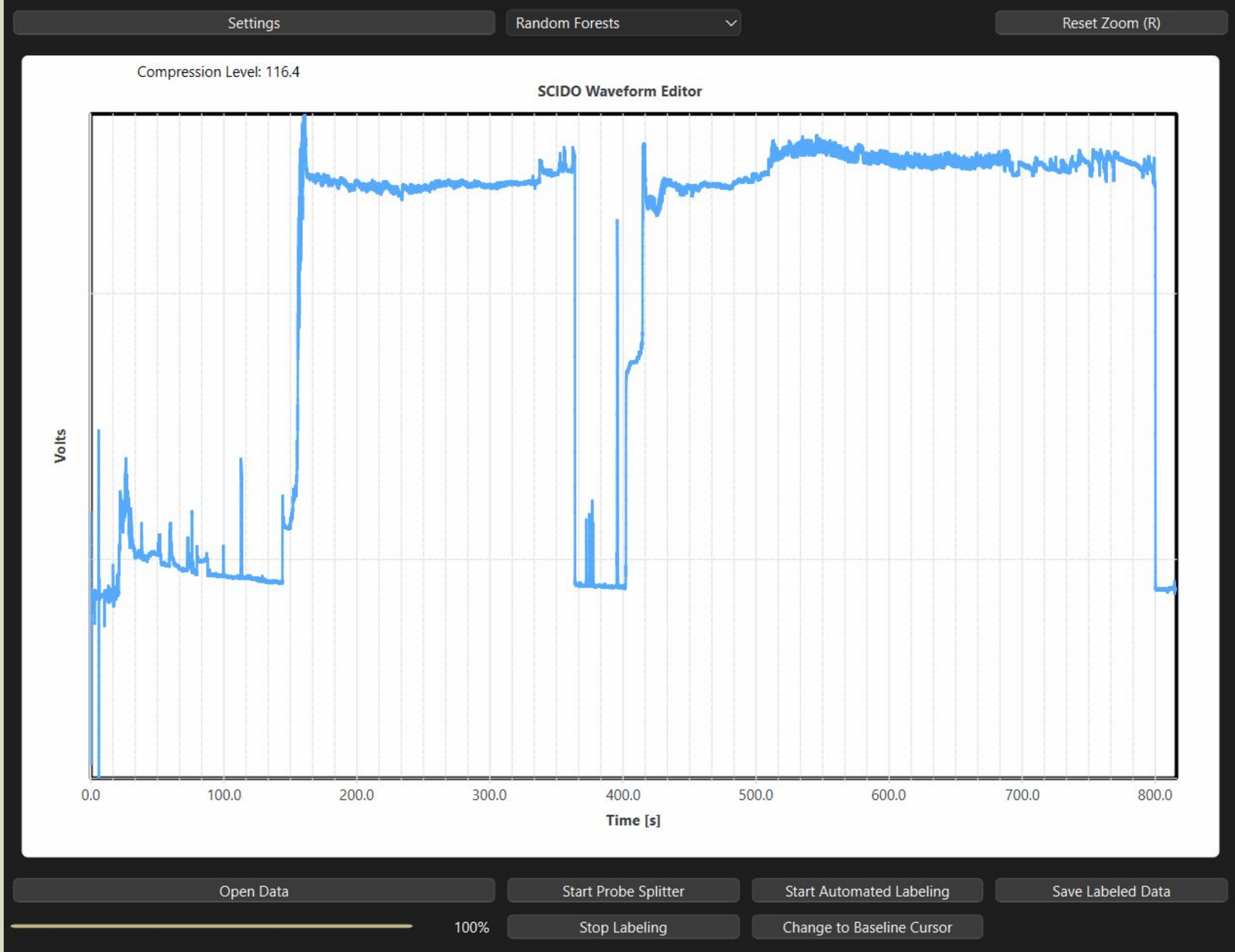
The Software

Supervised
Classification of
Insect
Data and
Observations

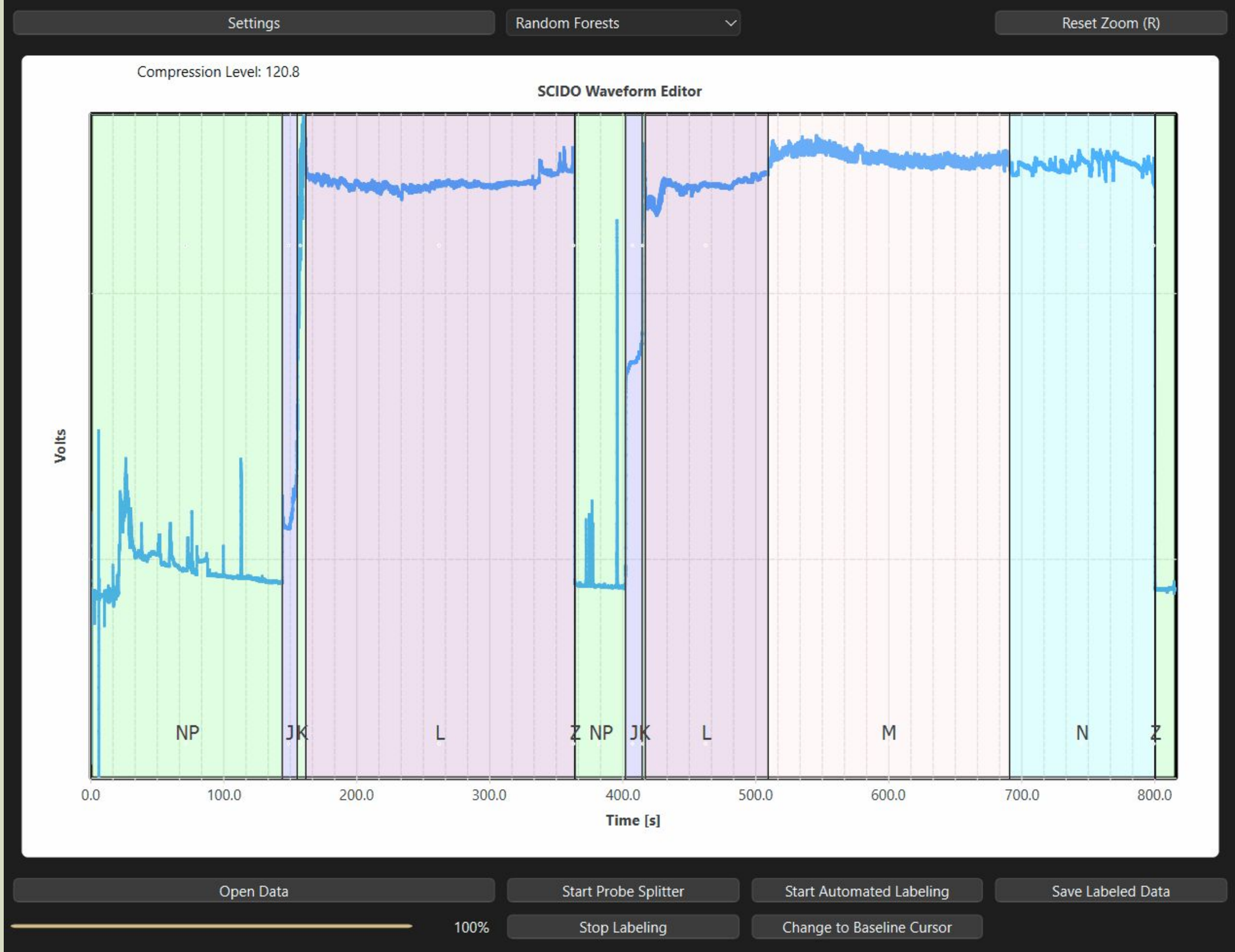


SCIDO

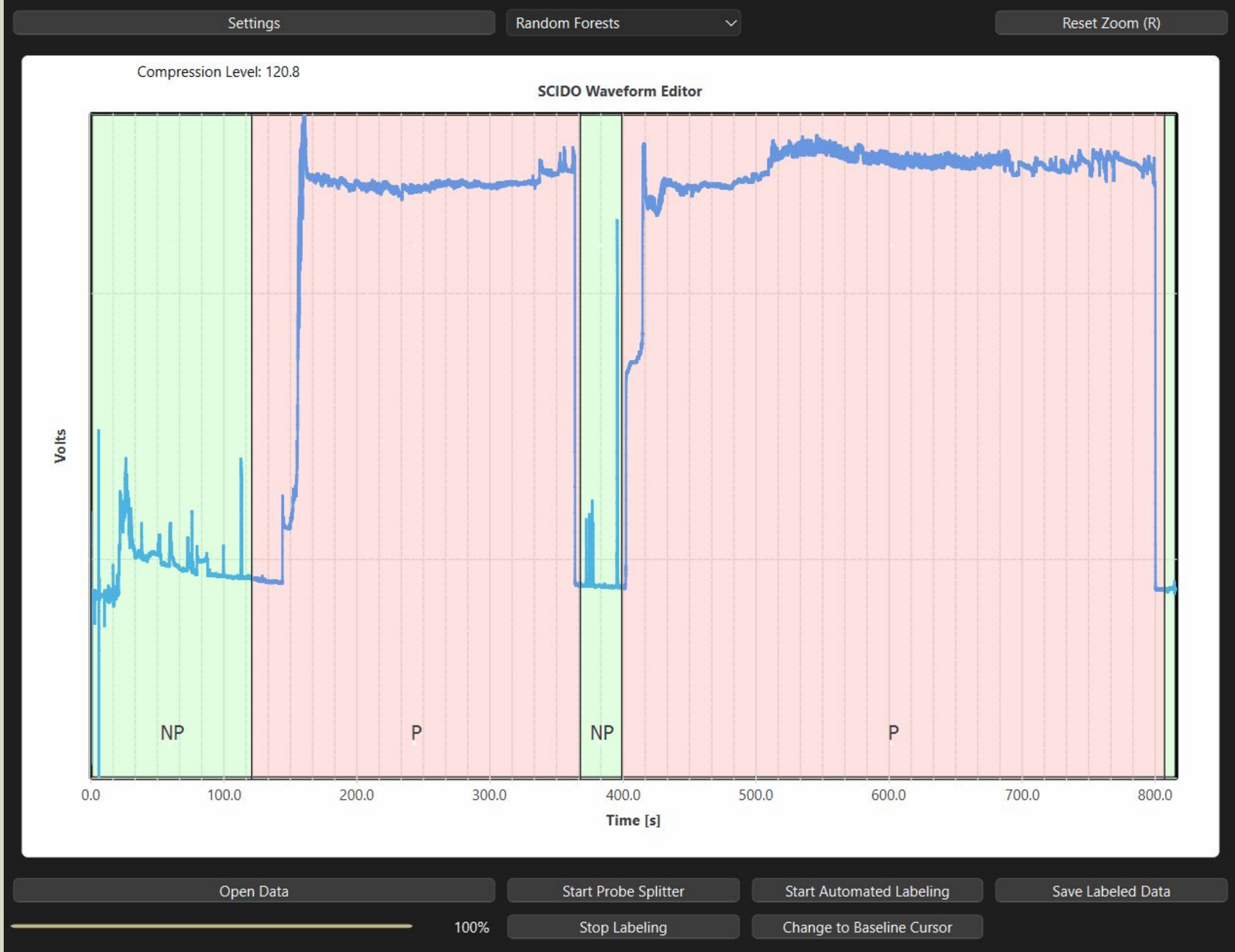
Opening Files



Probe Splitting



ML Labeling


Settings
Random Forests
Reset Zoom (R)
Compression Level: 120.8
SCIDO Waveform Editor
Volts
NP
P
NP
P
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
Time [s]
Open Data
Start Probe Splitter
Start Automated Labeling
Save Labeled Data
100%
Stop Labeling
Change to Baseline Cursor

Viewing & Editing

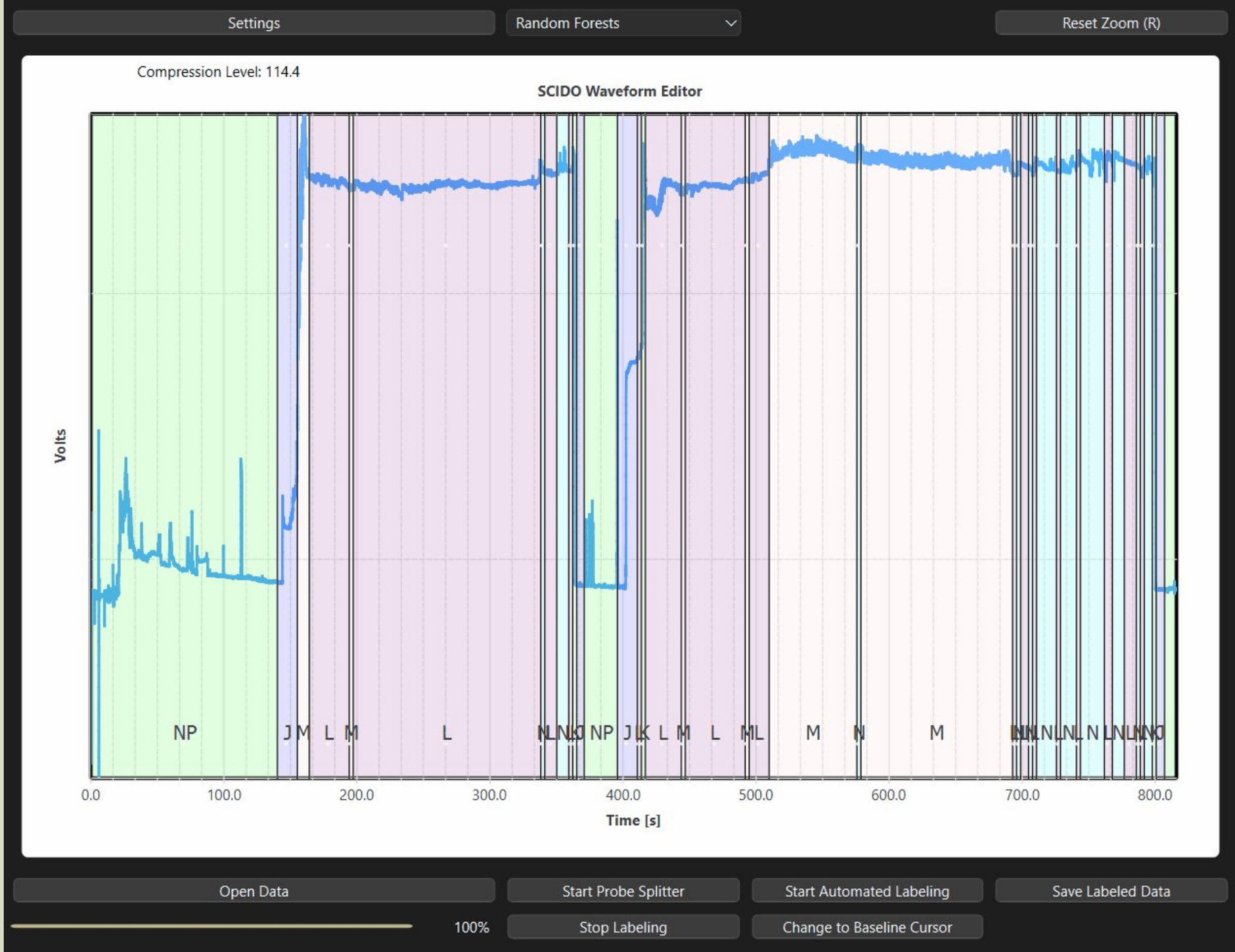
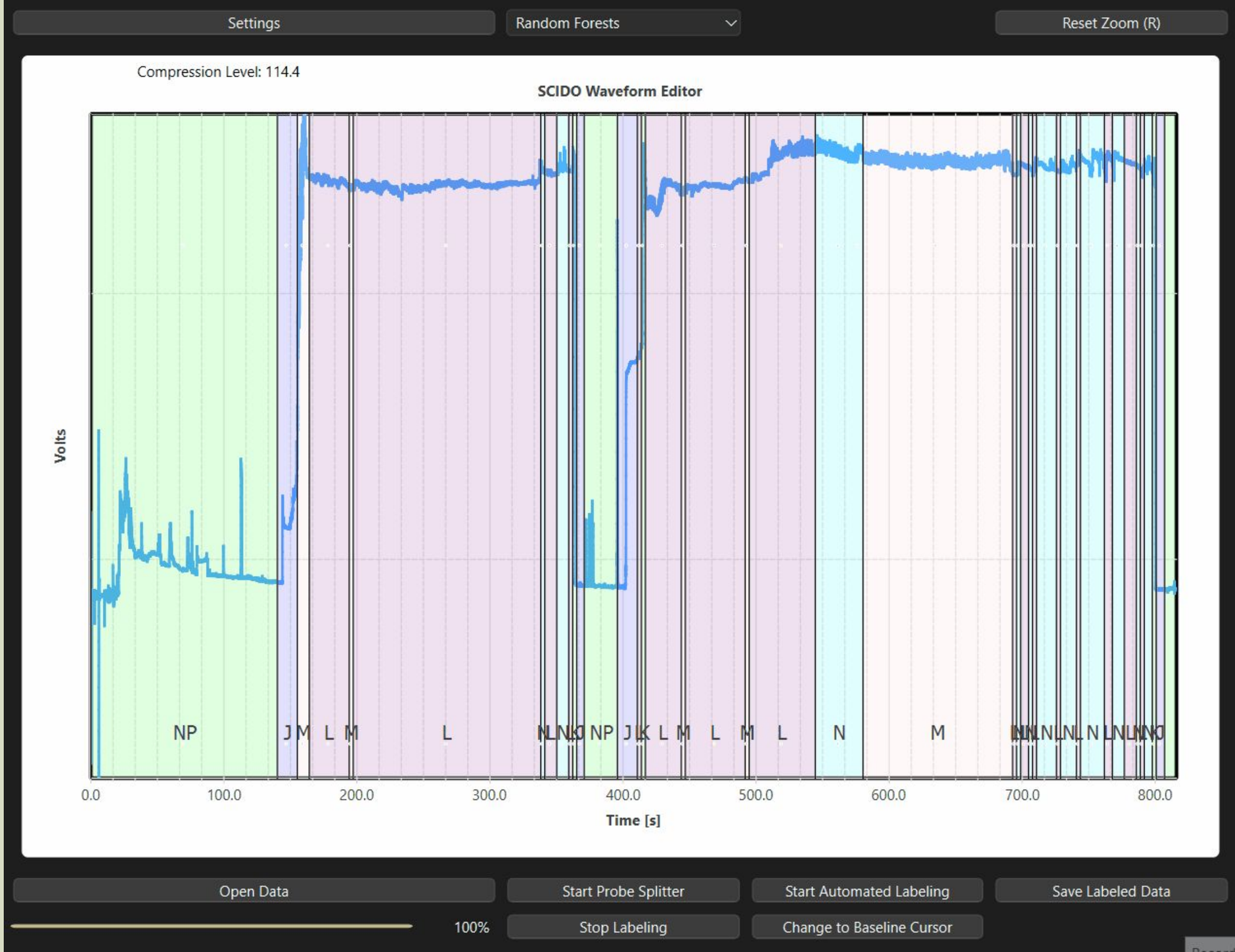
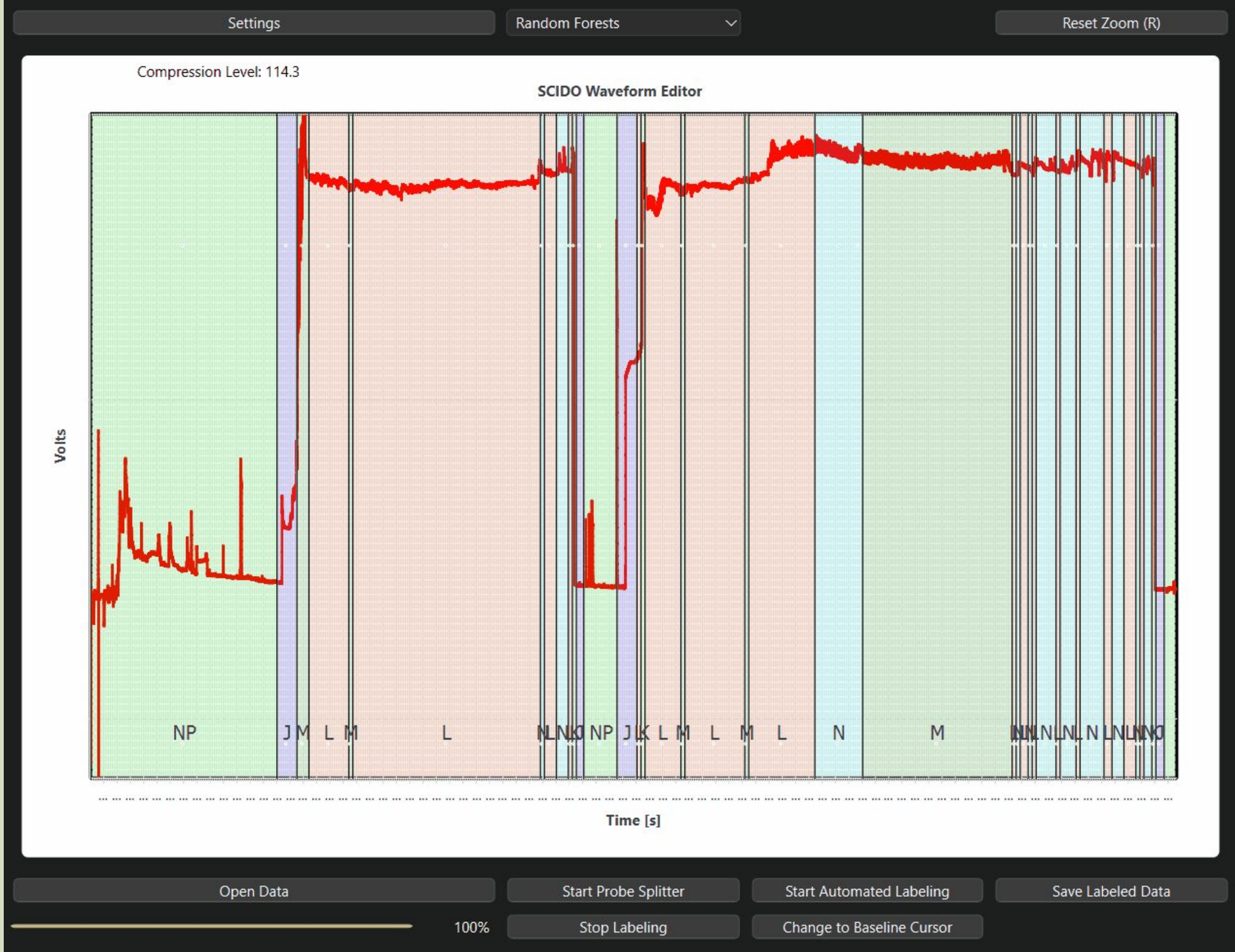


Chart Visuals



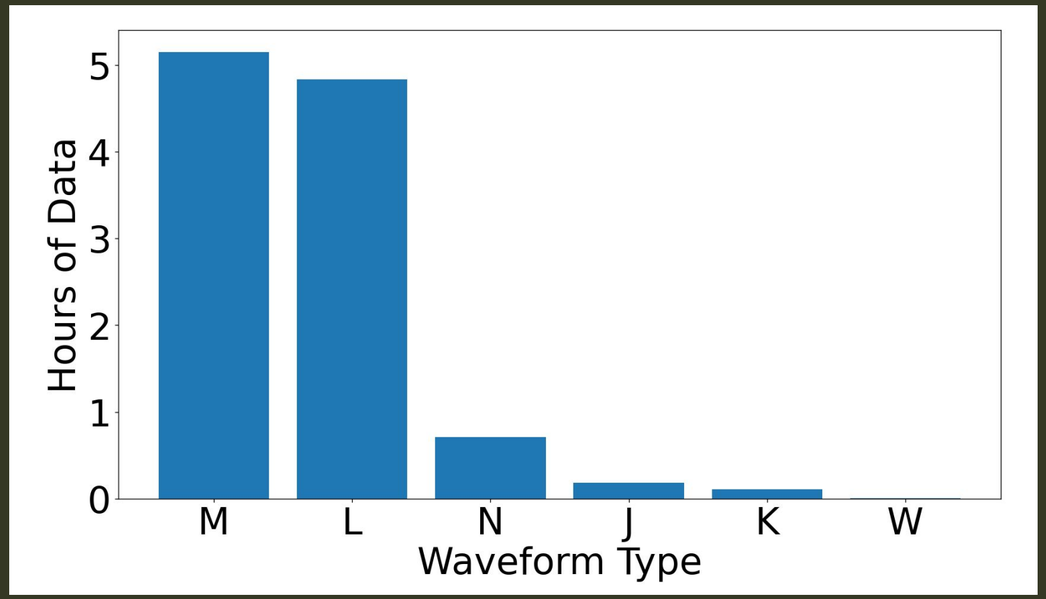
Saving Files



The Data

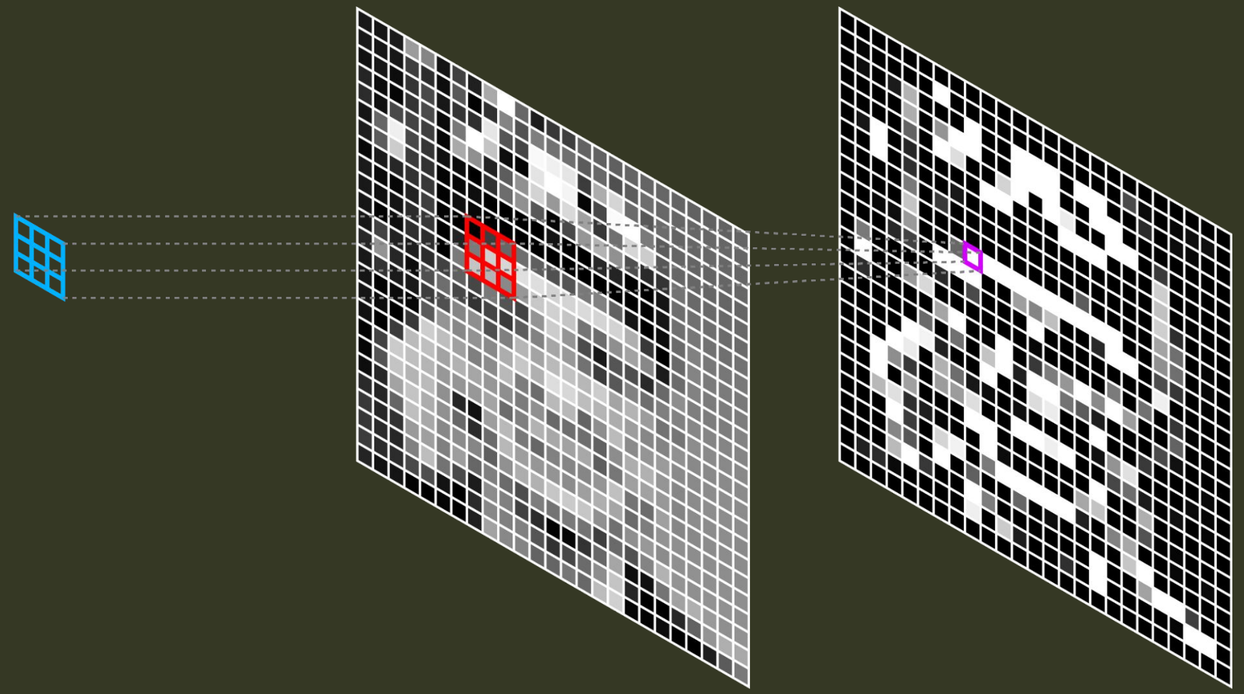
- Annotated EPG recordings
- 62 files
- 94 probes
- about 11 hours of probing data

Highly imbalanced data



L: searching for a blood vessel
M: sucking blood
W: withdrawal

Convolutions



-1	-1	-1
-1	8	-1
-1	-1	-1

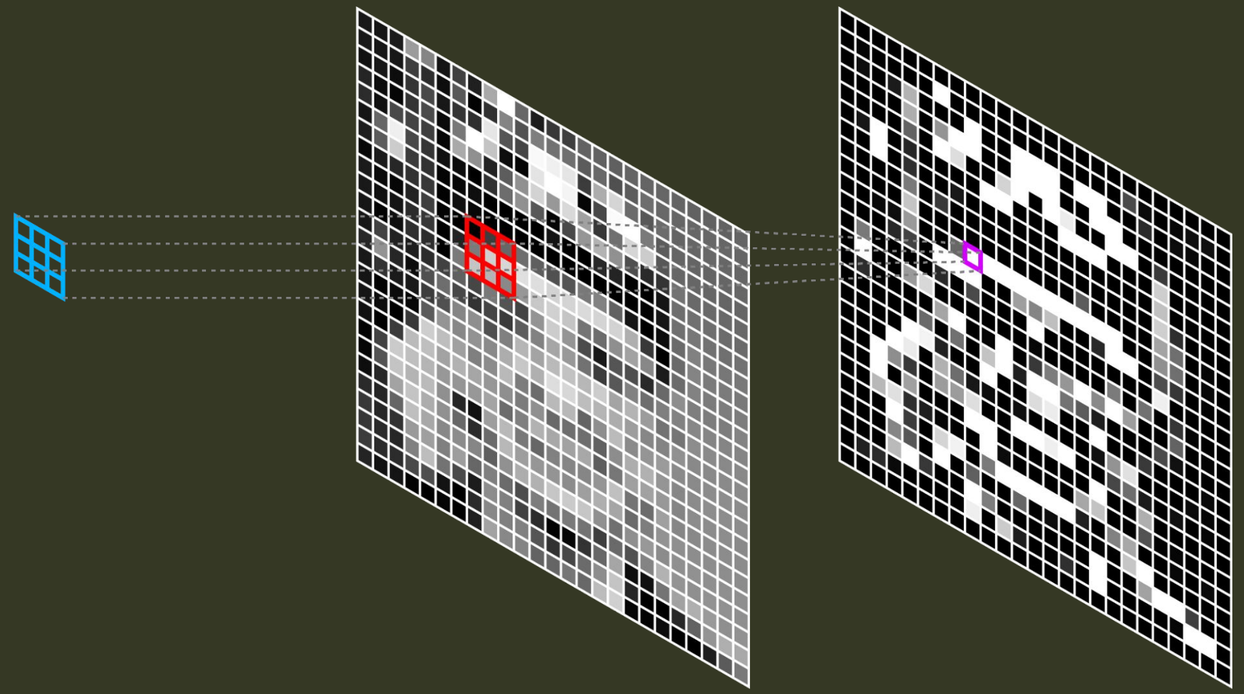
45	81	87
194	203	215
164	116	131

255*

* = max(255, 657)

Kernel **Input** **Output**

Convolutions



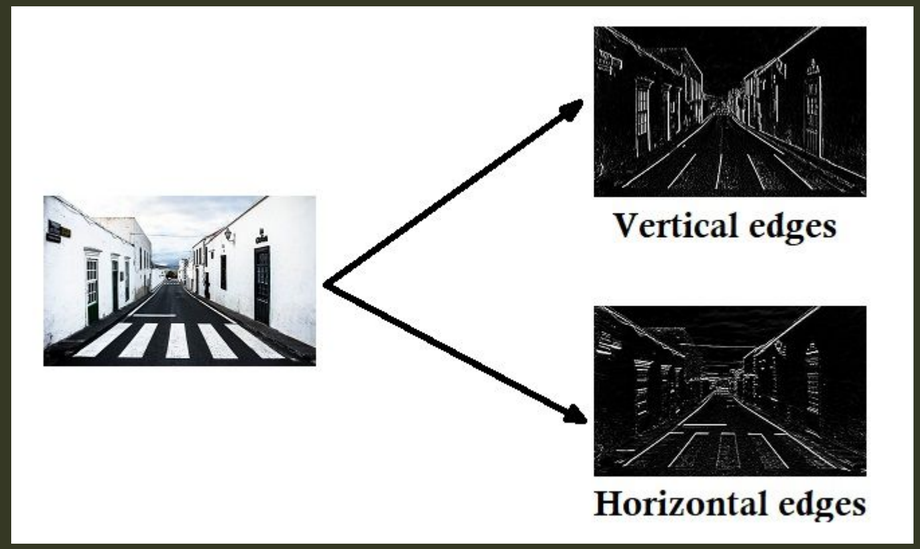
-1	-1	-1
-1	8	-1
-1	-1	-1

45	81	87
194	203	215
164	116	131

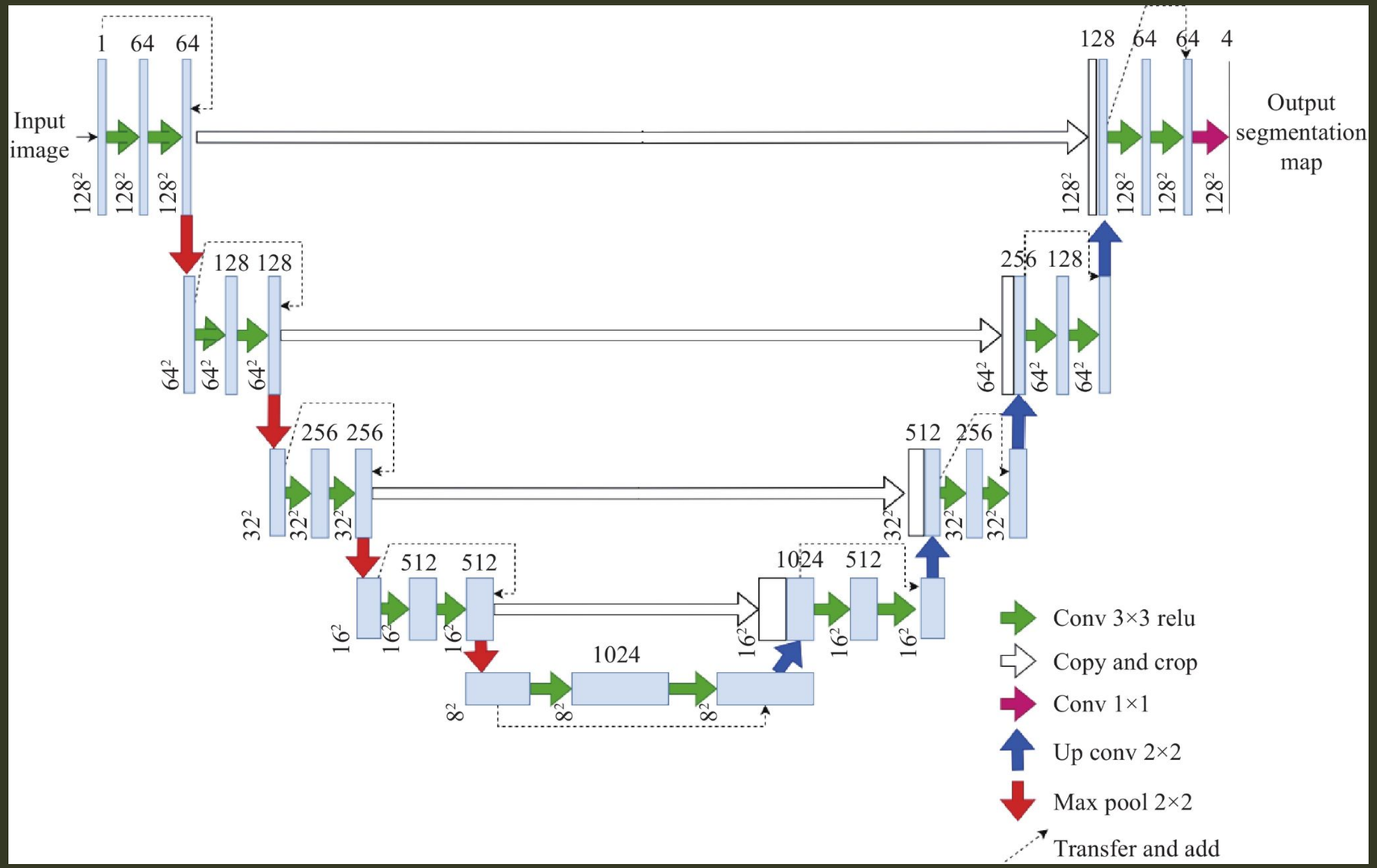
255*

* = max(255, 657)

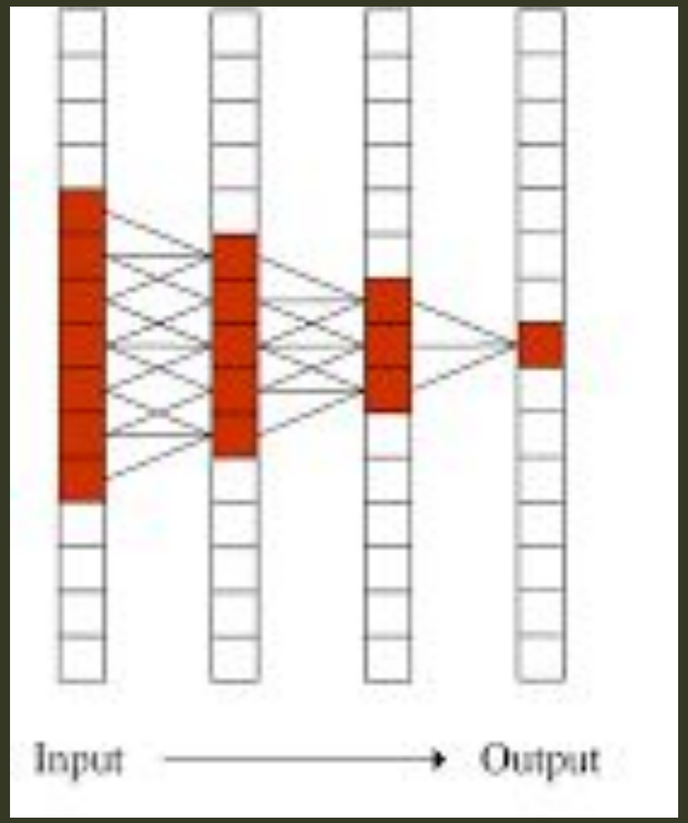
Kernel **Input** **Output**



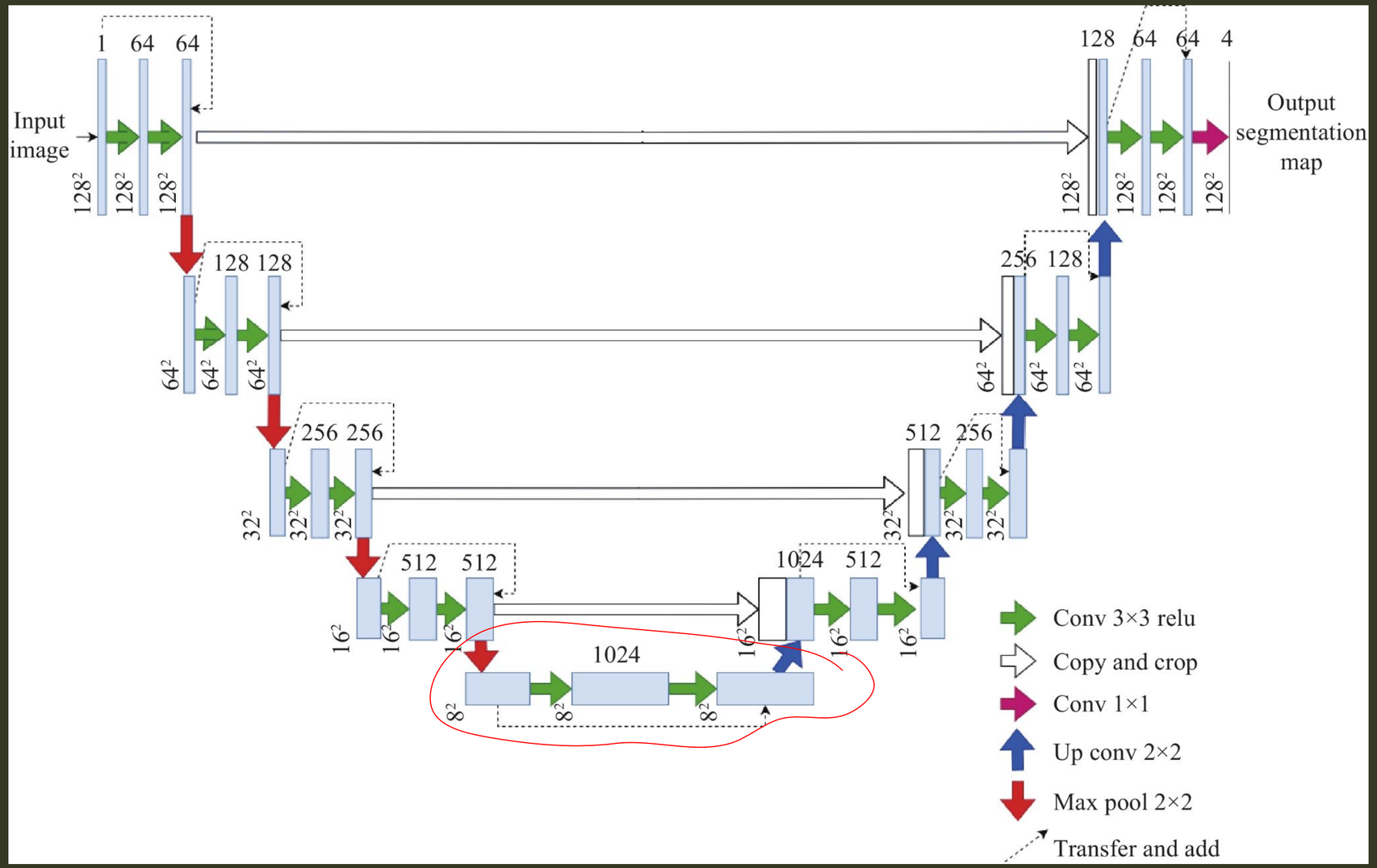
What is a UNet



A Problem: Receptive Field



Adding Attention



The Best Model: UNet (+ Attention)

Pros:

- Can see a lot of the data at once (large receptive field)
- Powerful

Cons:

- Can memorize the training data (overfitting)

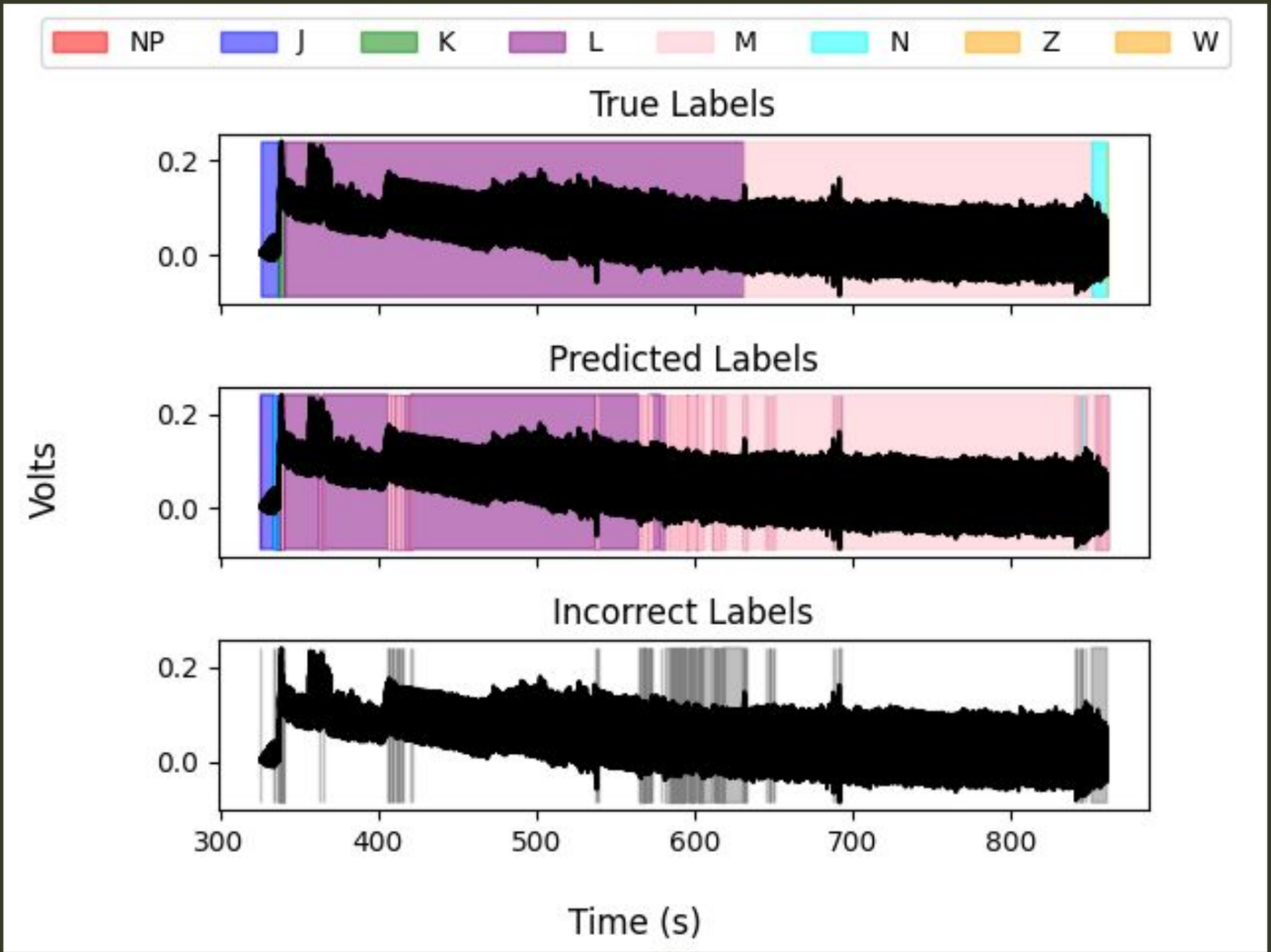
The Best Model: UNet (+ Attention)

Accuracy: ~82%

F1 score by waveform type:

	J	K	L	M	N	W
UNet	0.88	0.67	0.83	0.87	0.32	0.62

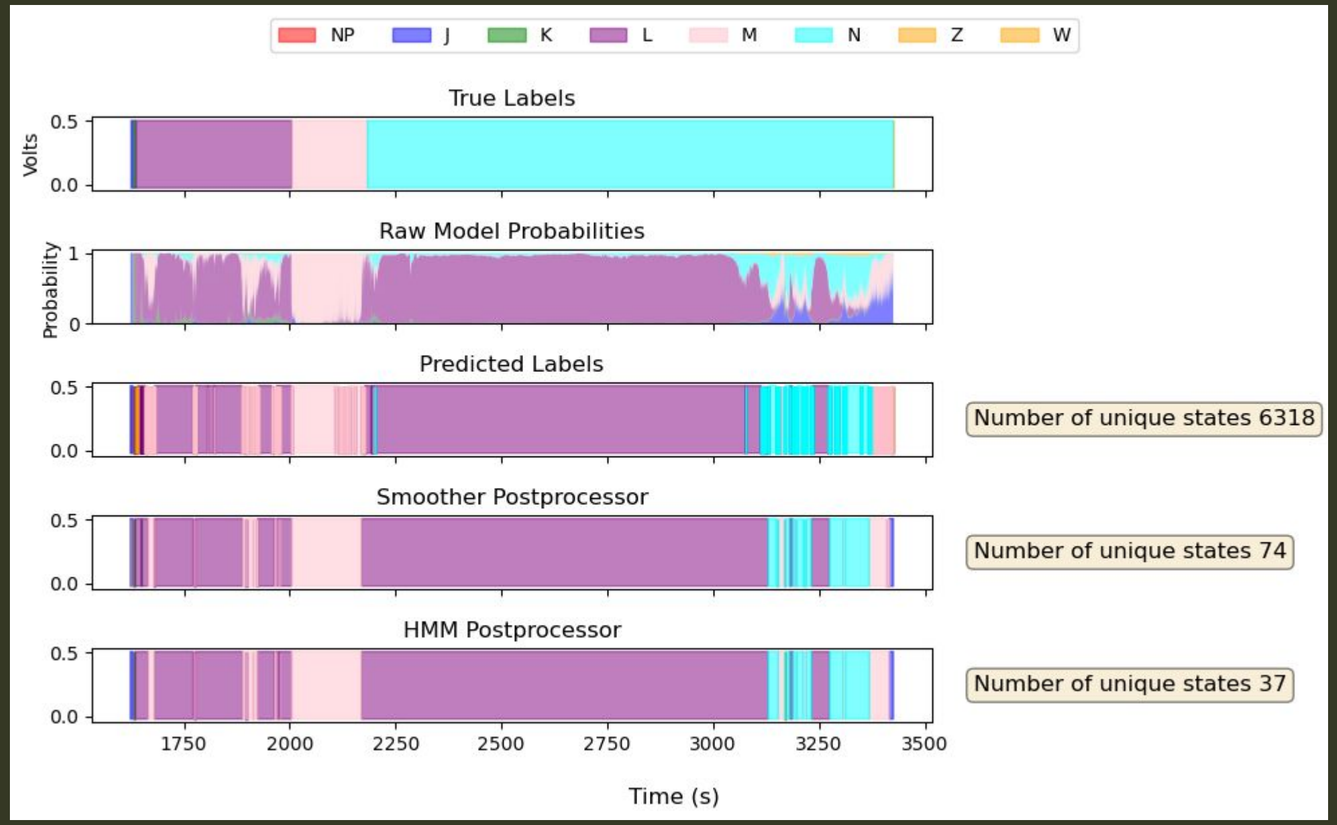
The Best Model: UNet (+ Attention)



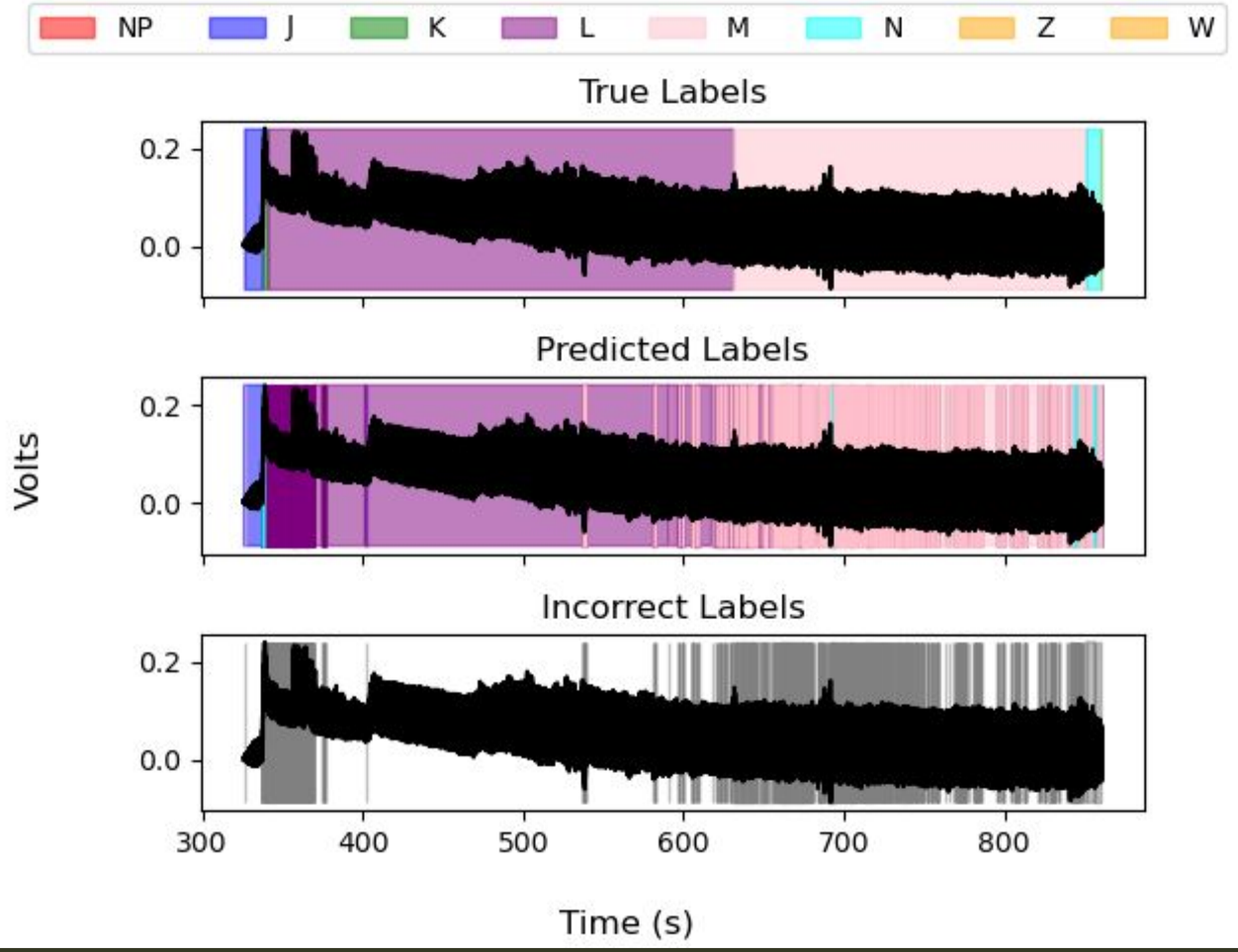
Observe barcodes

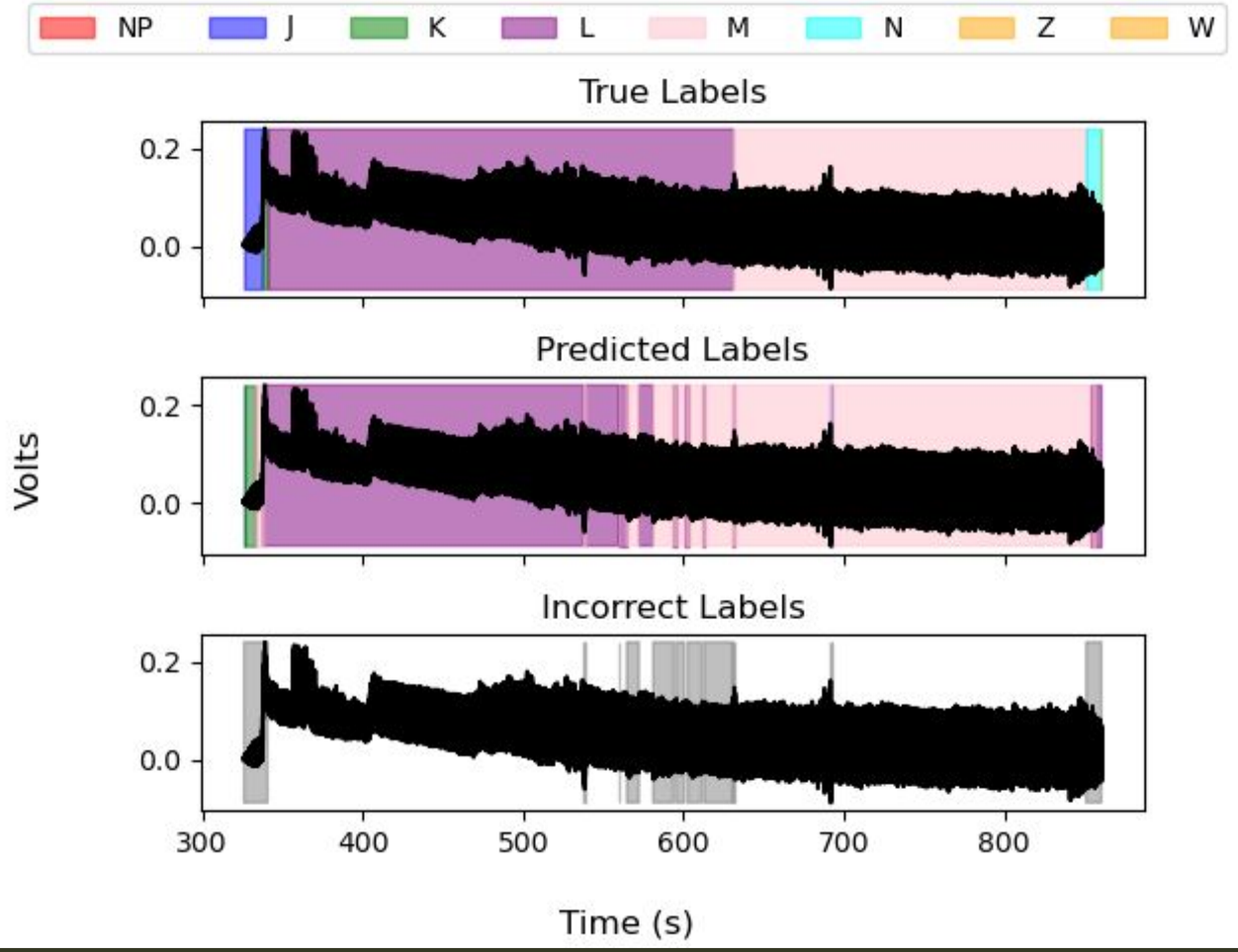
Post Processing

- Problem: “Barcodes”
- Solutions
  - Smoothing filter
  - Barcode cutter
  - HMM
  - HSMM
- Project for next interns



Example: HMM Postprocessor



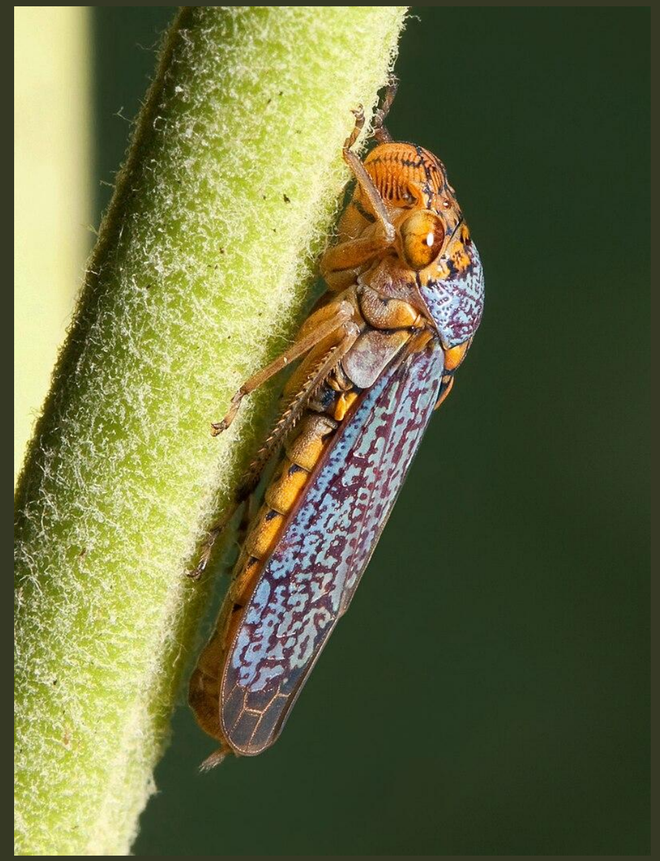


Limitations

- Liaisons are experts in biology but not computer science
- Model output is never perfect, needs some manual adjustment
- Important to convey system limitations
  - Potential for negative impact on science if blindly trusted

Future Work

- User Interface
  - Make more comprehensive by integrating live view
  - Render long recordings more efficiently
  - Integrate post-processing
- Machine Learning
  - Other post-processing and data augmentation methods
  - Adapt models for other insects



Thank you!

Questions/comments/suggestions?

We’d love to touch on ML models we’ve tried or the project’s future!

Acknowledgements
USDA (58-2034-3-445)
USDA (58-3022-4-034)
NSF (DBI - 2304787)