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

Aphid

Sharpshooter

Mosquito

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

Aphid

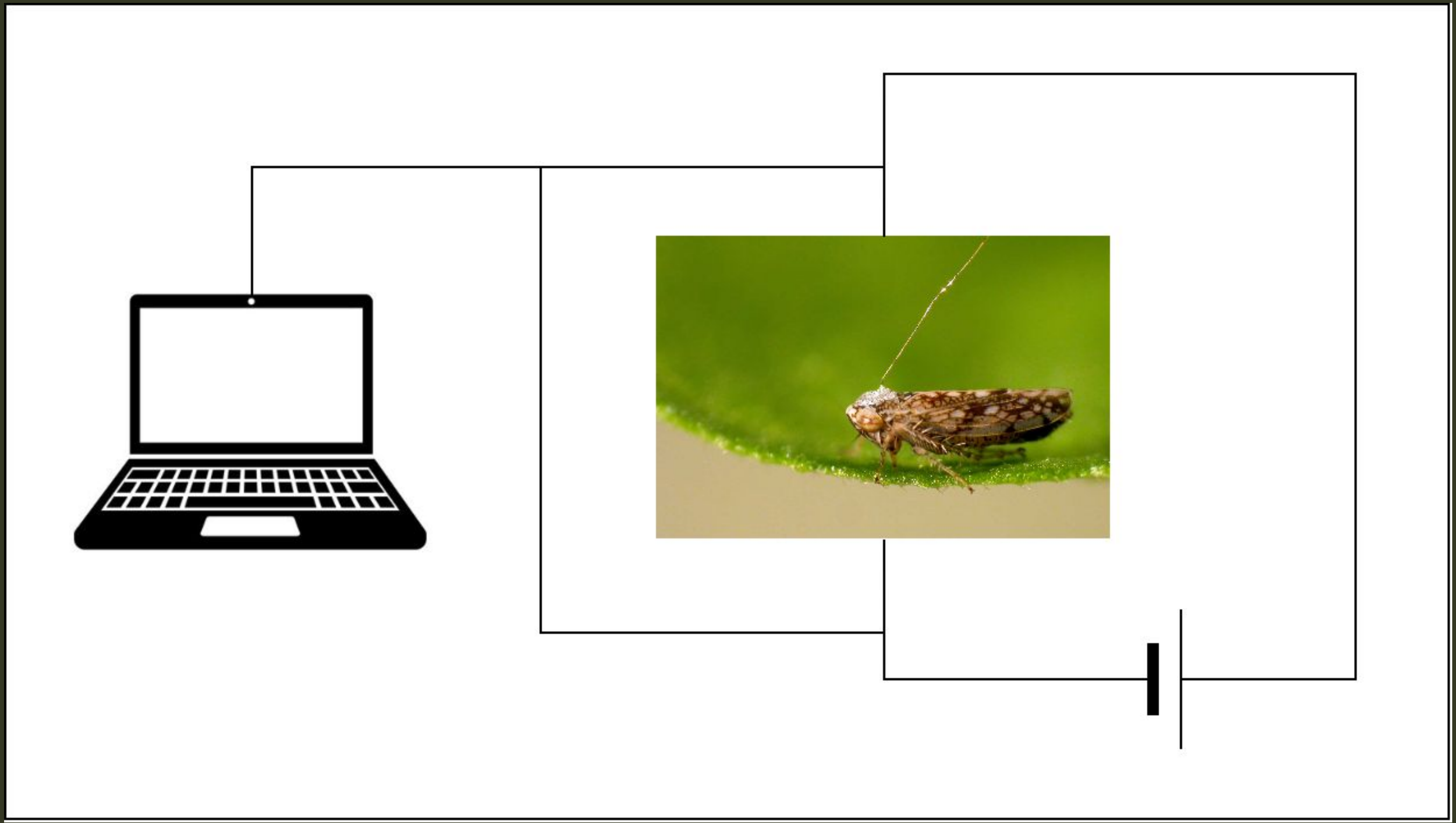
Sharpshooter

Mosquito

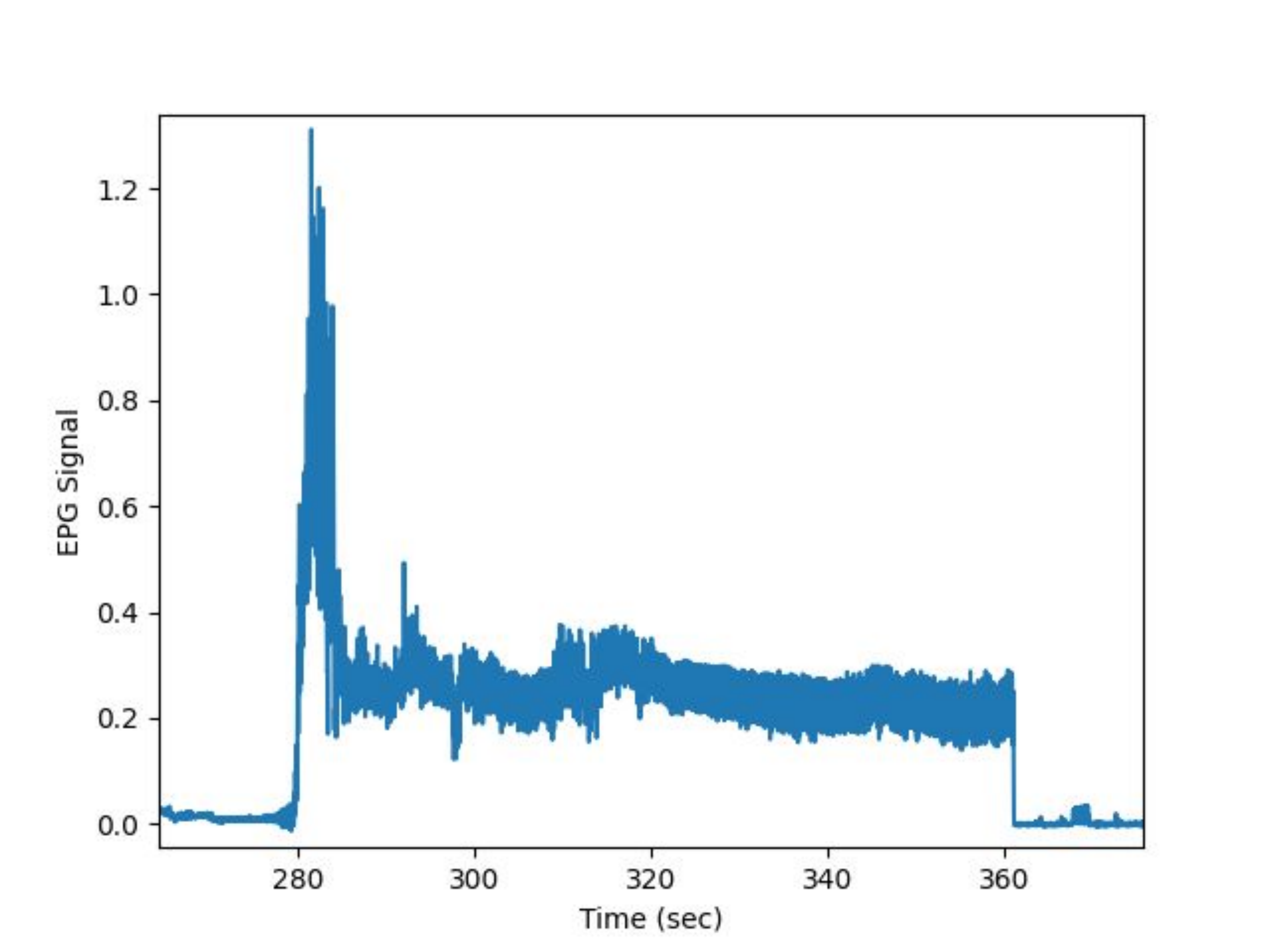
We can’t directly observe what the mouthparts are doing

Electropenetrography (EPG)

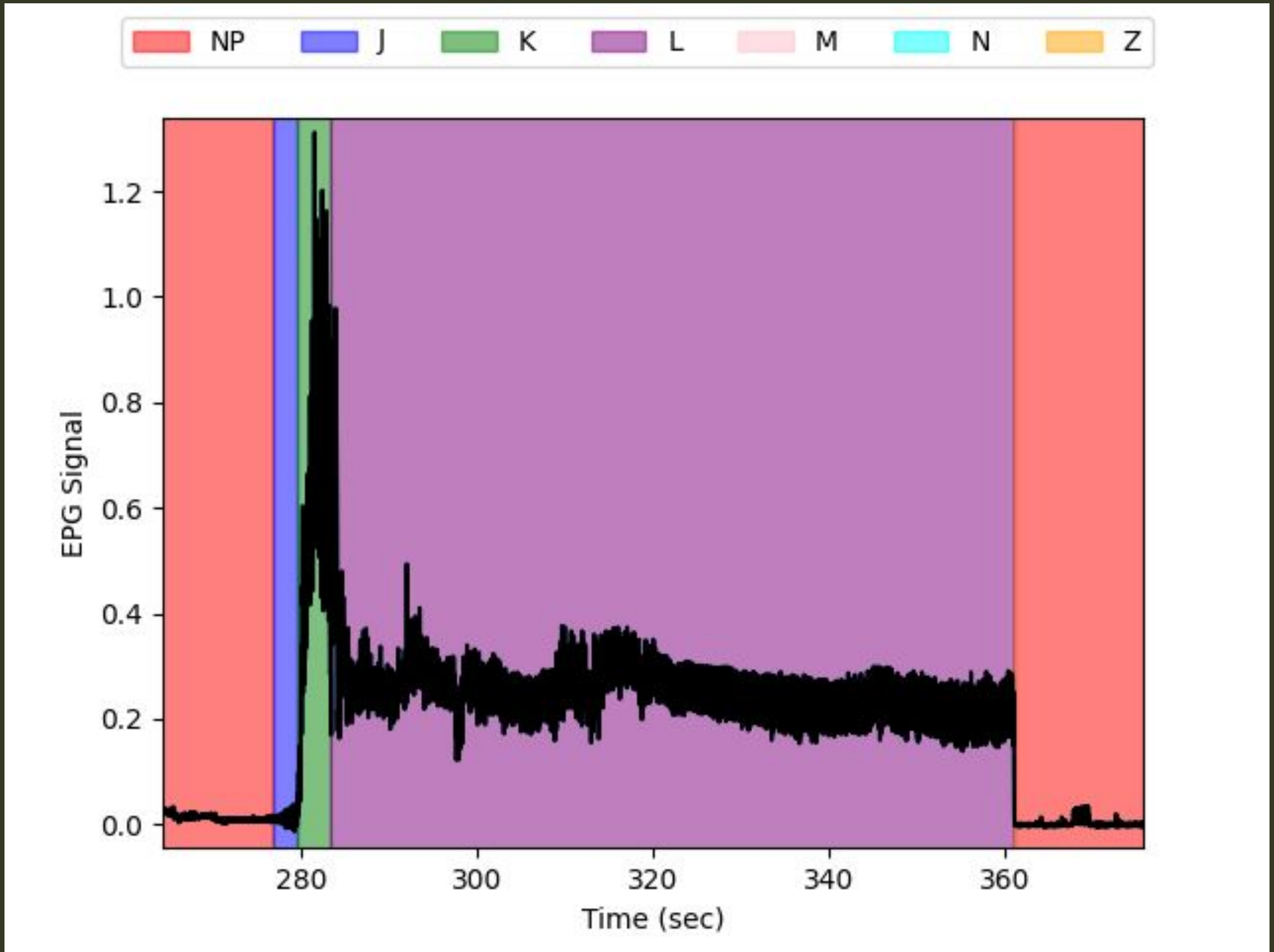
Leafhopper ready for EPG



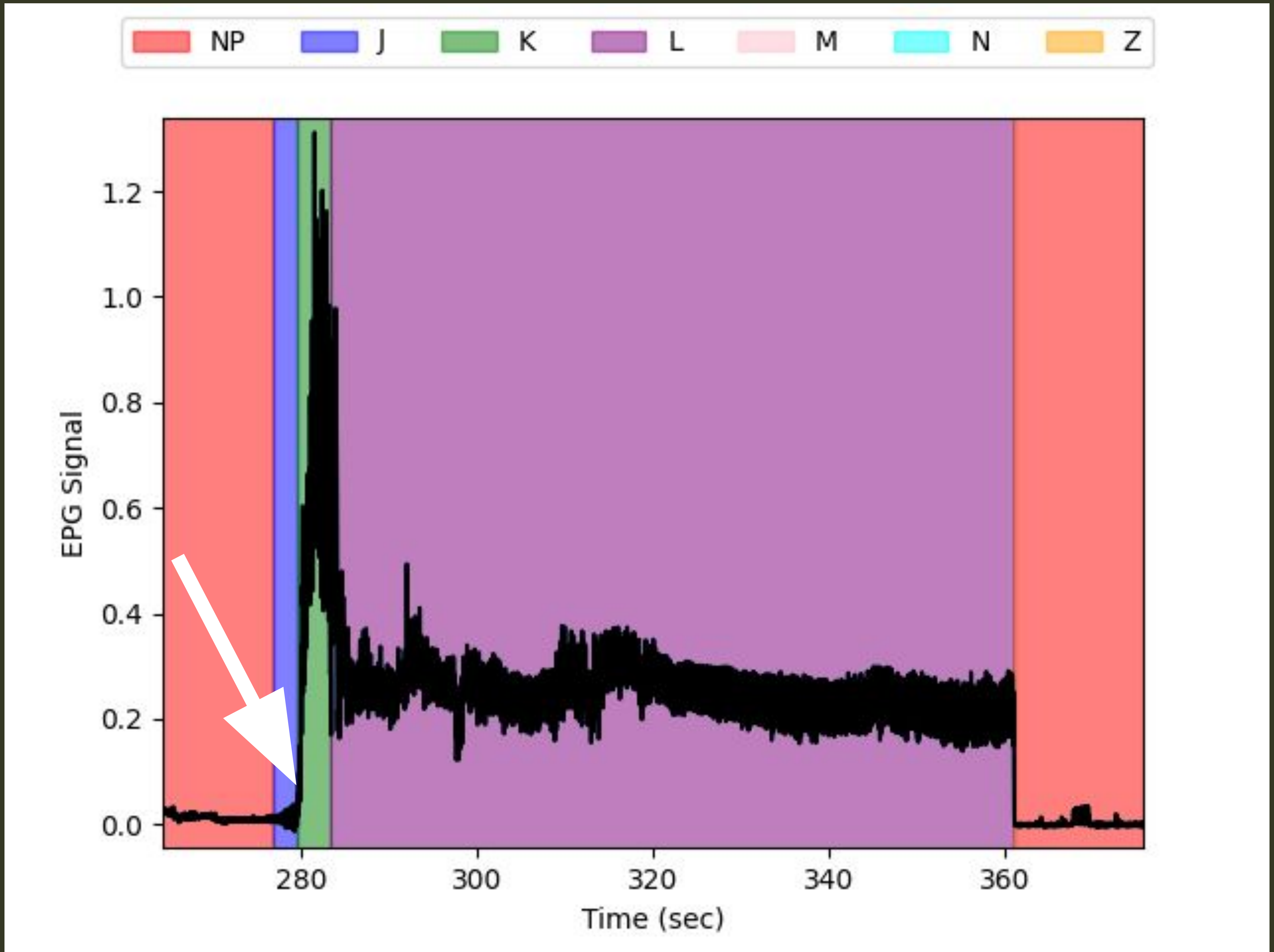
EPG Circuit



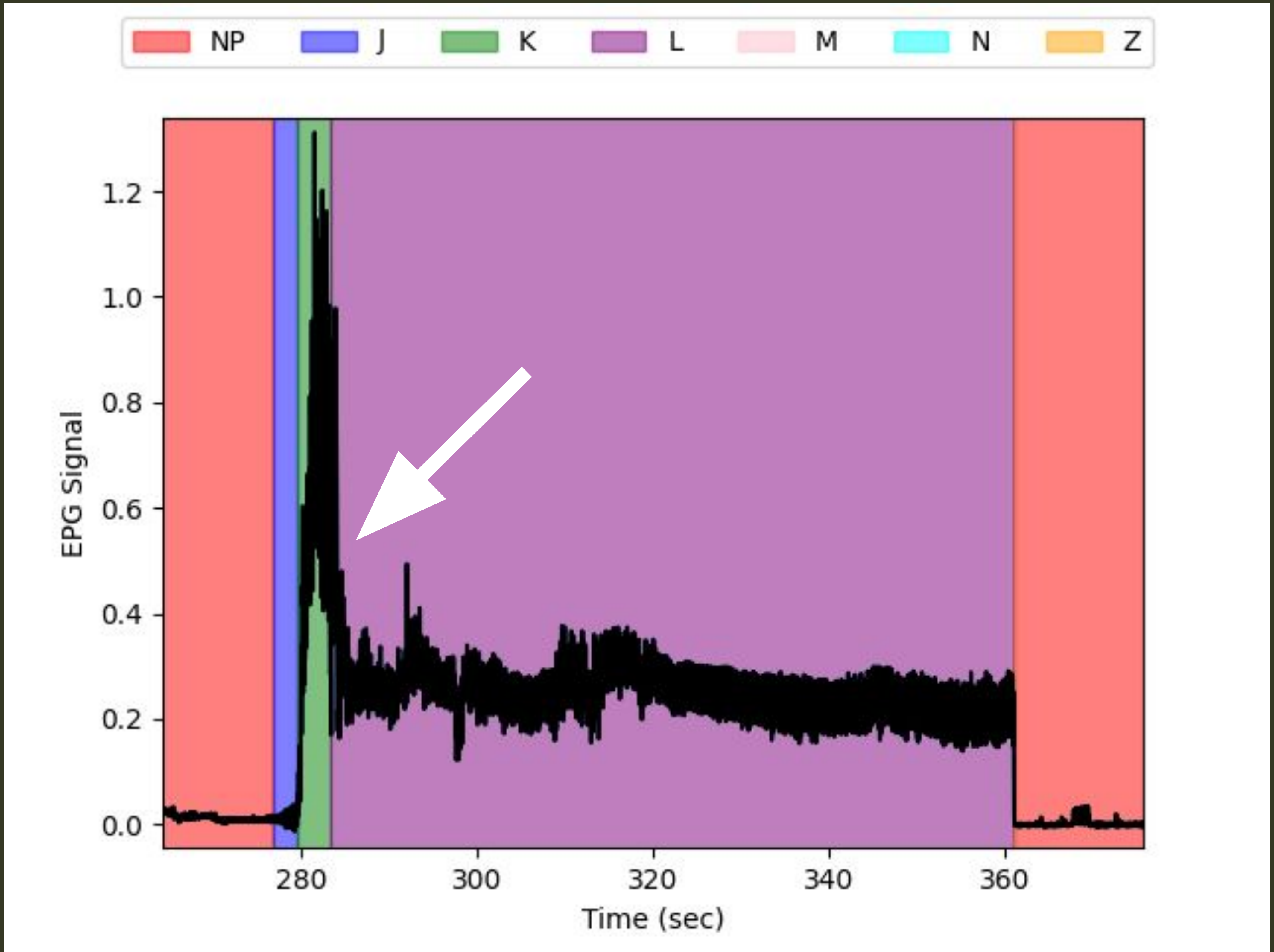
EPG Recording



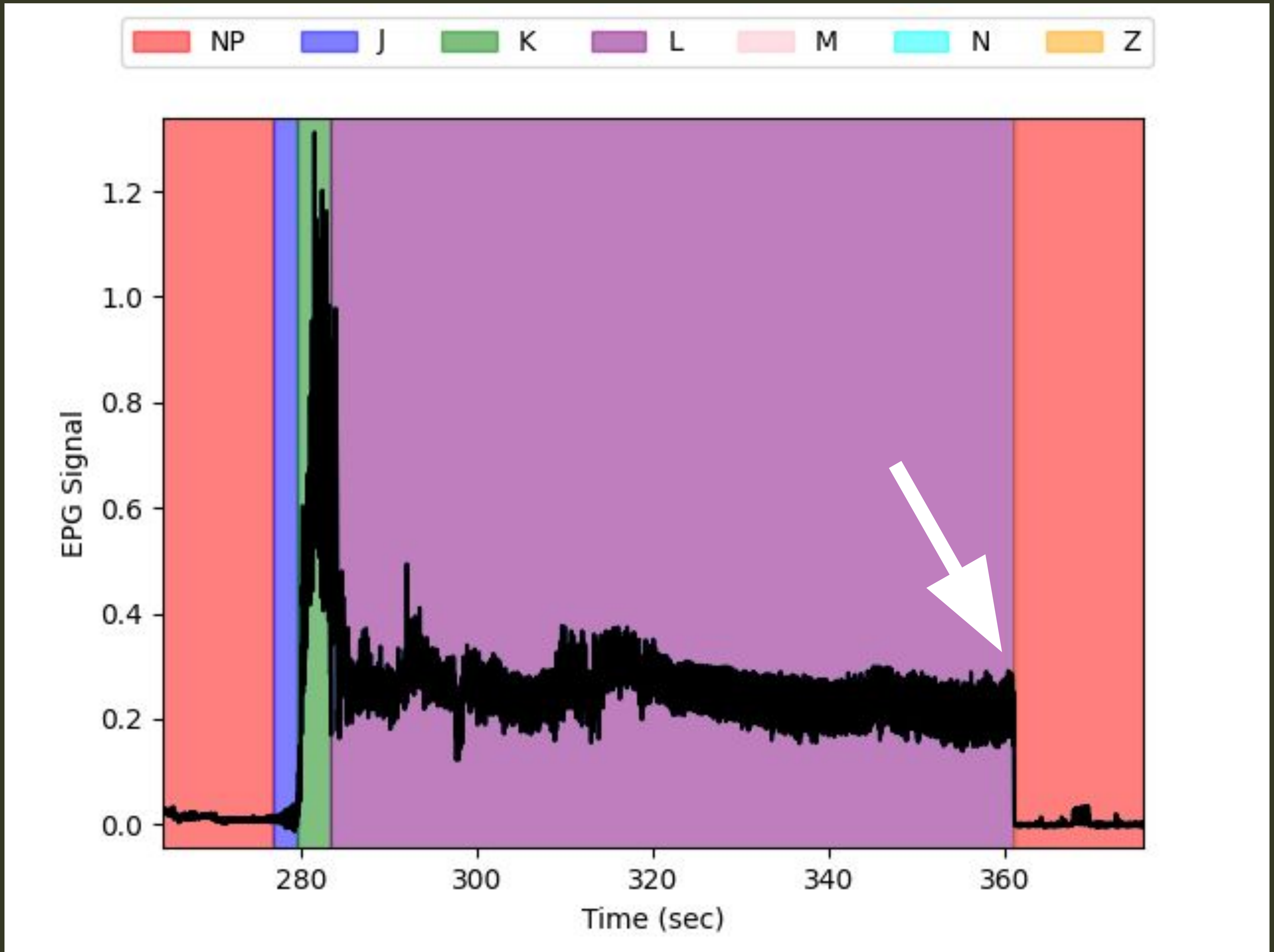
EPG Recording



EPG Recording



EPG Recording



EPG Recording

Our task: Automate EPG labelling and make it accessible

Deliverables

Train predictive ML model(s) for waveform recognition

Present it with a user interface

Success Criteria

Machine Learning Goals

- Accurately label EPG recordings
- Integrate seamlessly with GUI

User Experience Goals

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

*Success = A model that is **intuitive** for scientists*

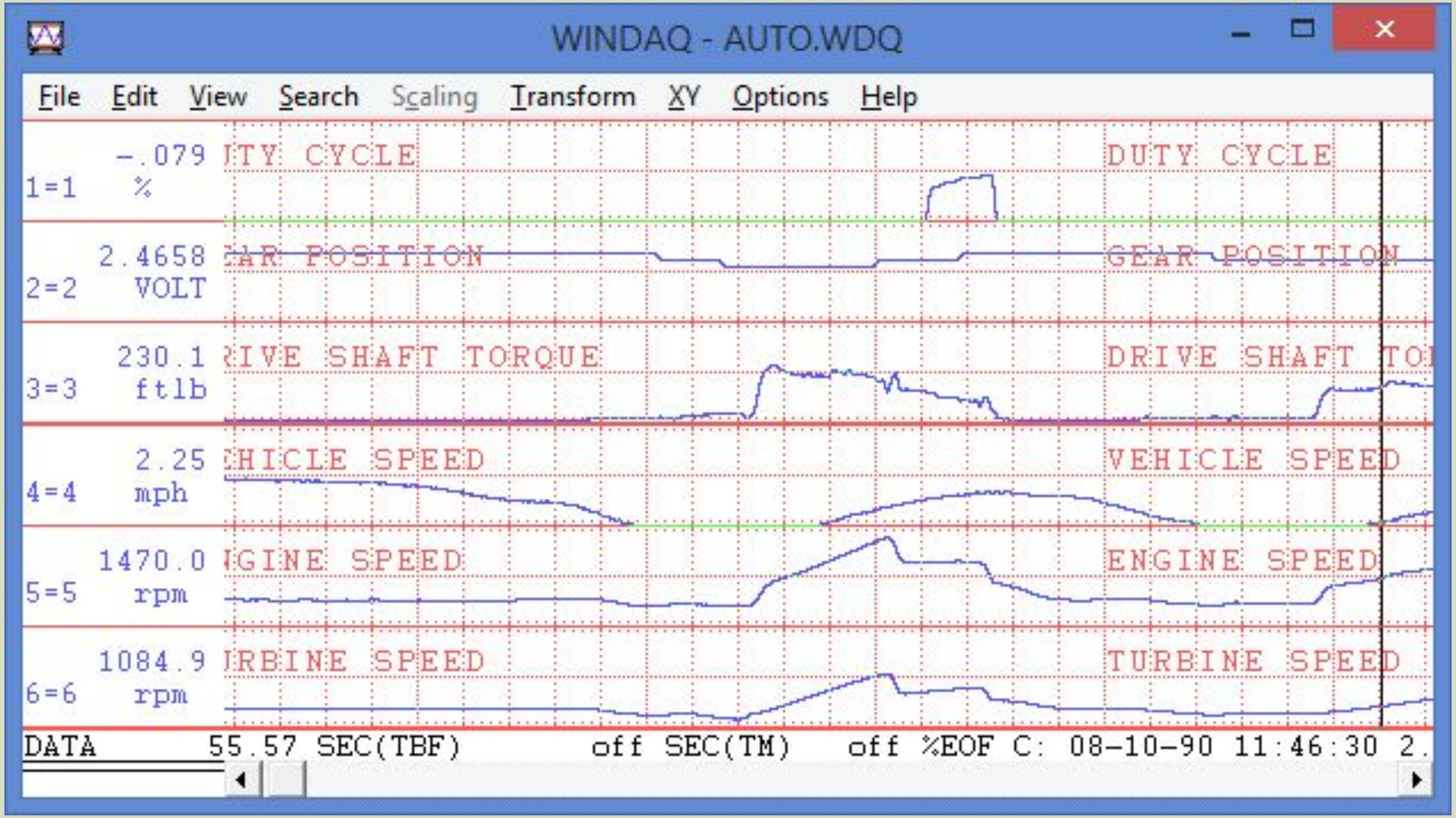
Why ML?

- Labeling is tedious for humans
- Not deterministic - no single algorithm
- Makes it perfect for a ML model!

- Automated recognition
- Removes human error

Why do they need a GUI?

- Windaq is inefficient and cumbersome
- Doesn't work with ML



GUI

Visualization (data-to-user)

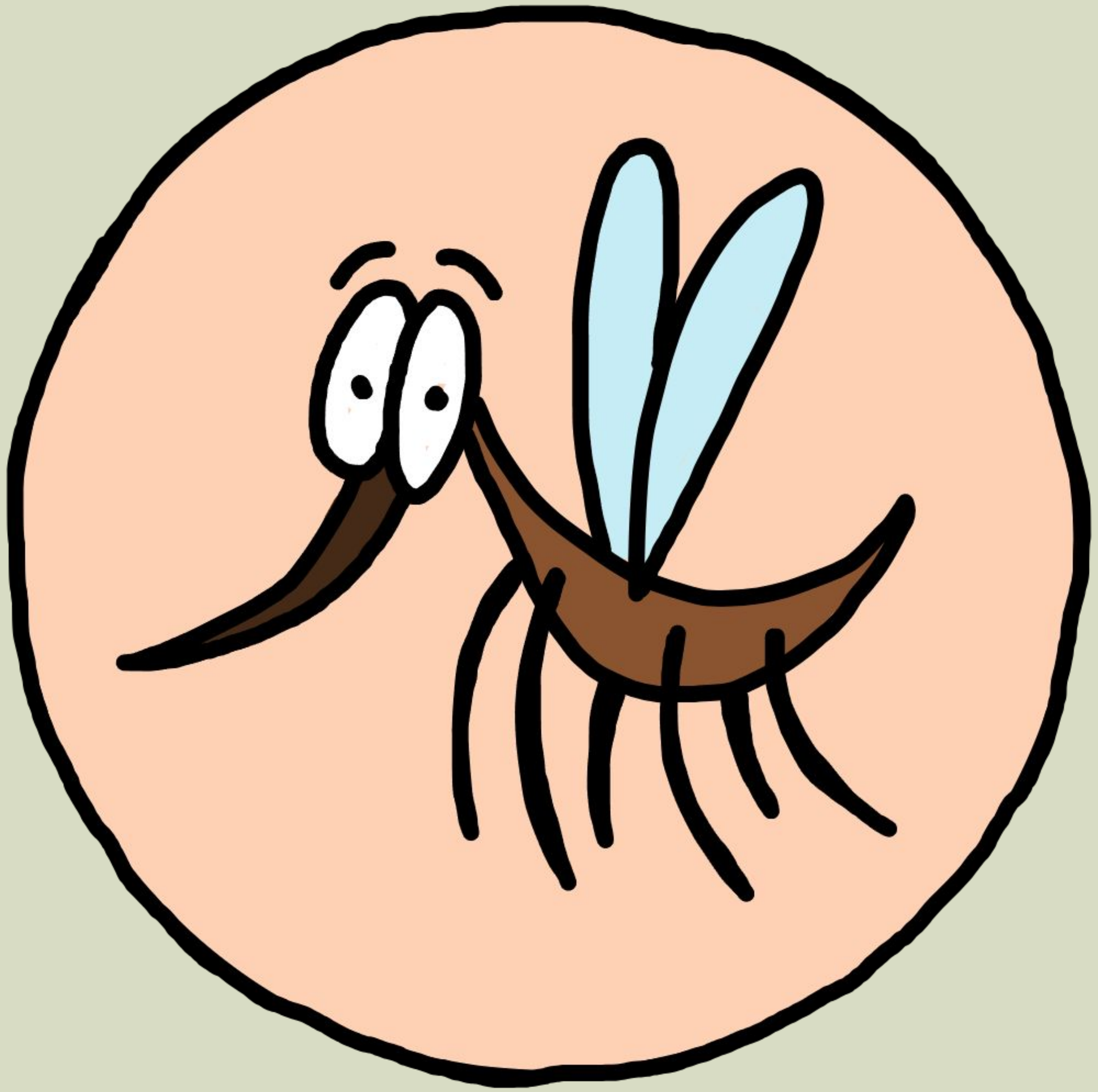
- Labeled EPG data in time series
- Color-coded regions highlighted
- Overall modernized experience compared to Windaq

Characterization (user-to-data)

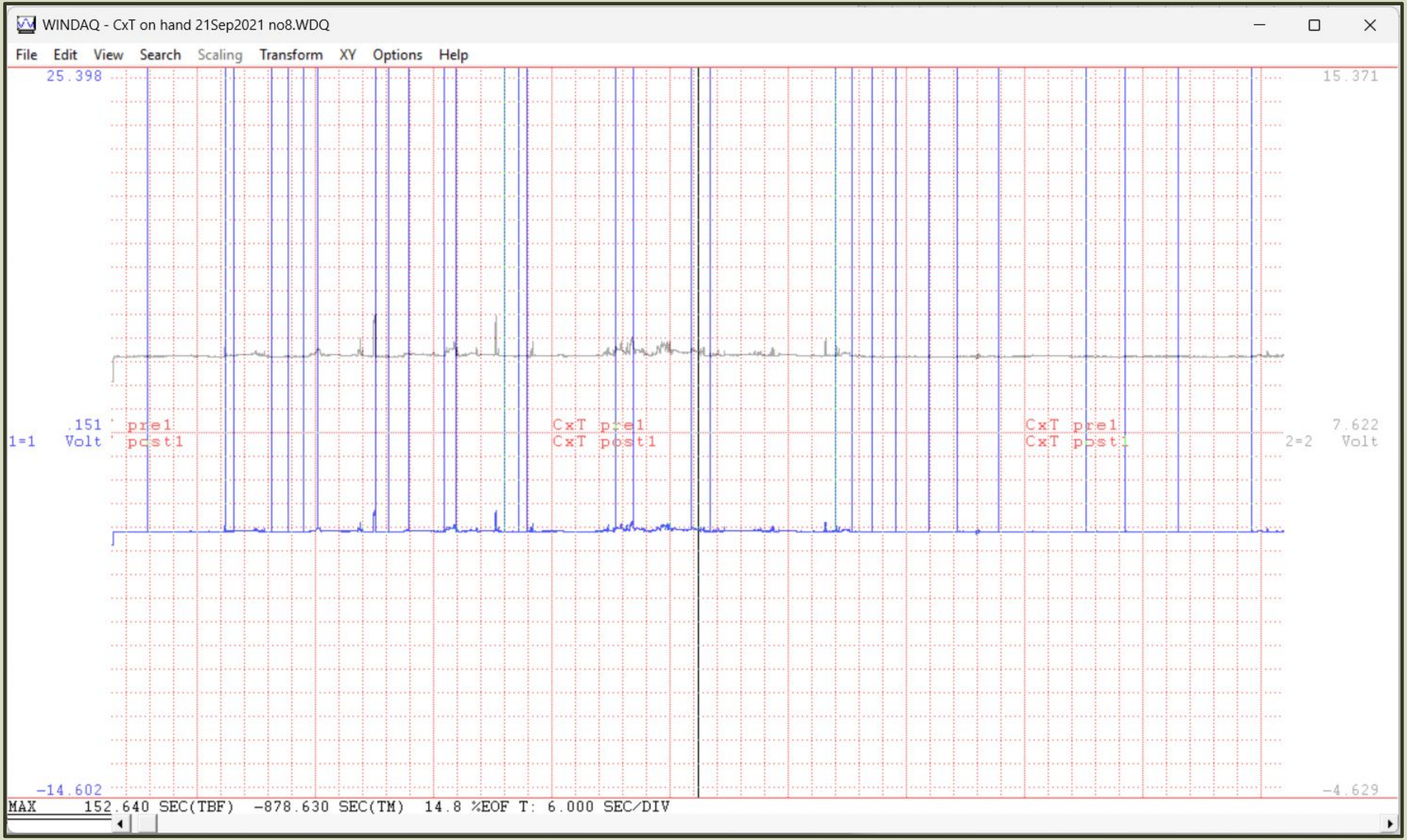
- Apply the ML model to data
- Adjust, delete, modify labels
- Characterization without alterations to dataset

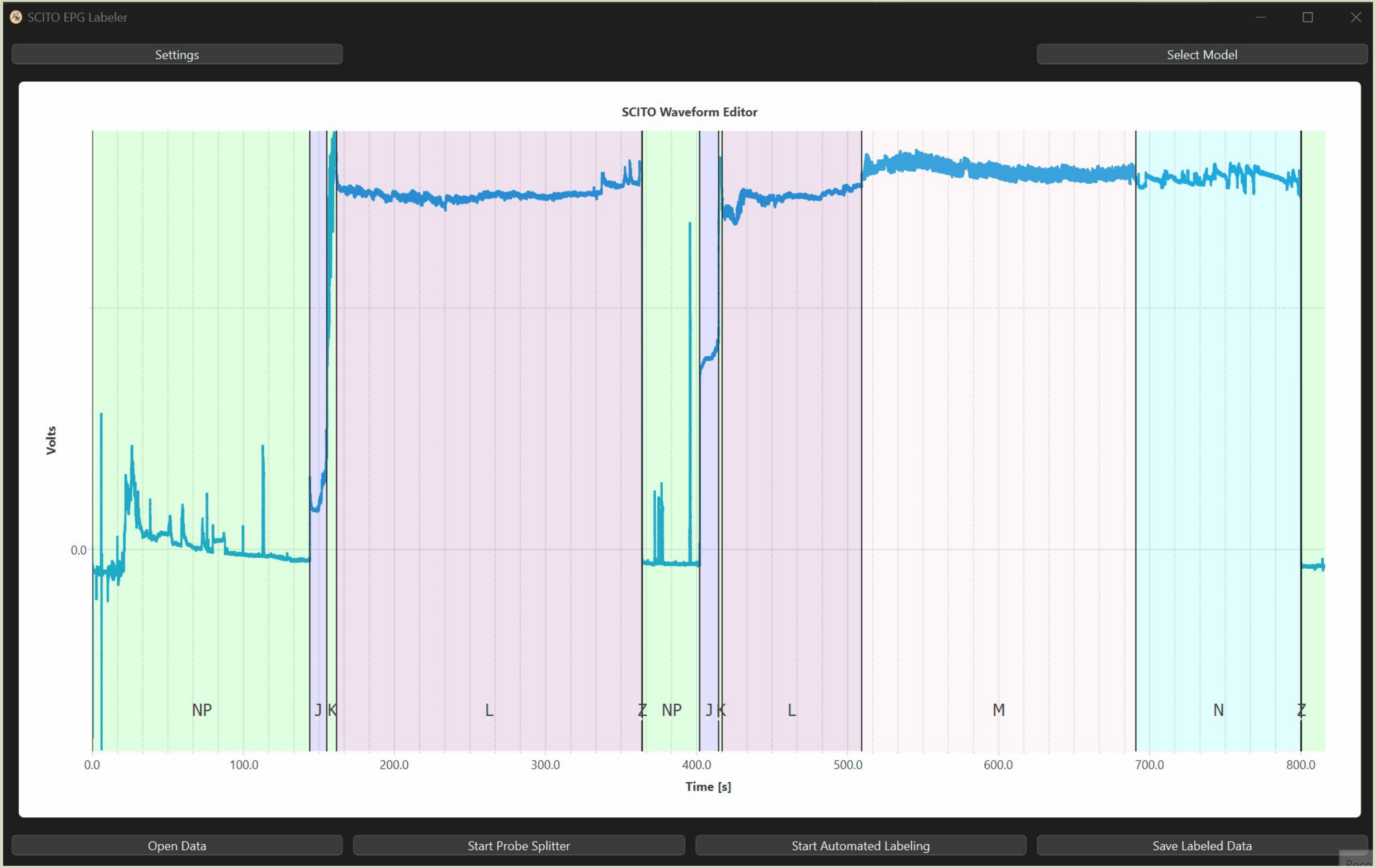
The Software

Supervised
Classification of
Insect
Data and
Outcomes

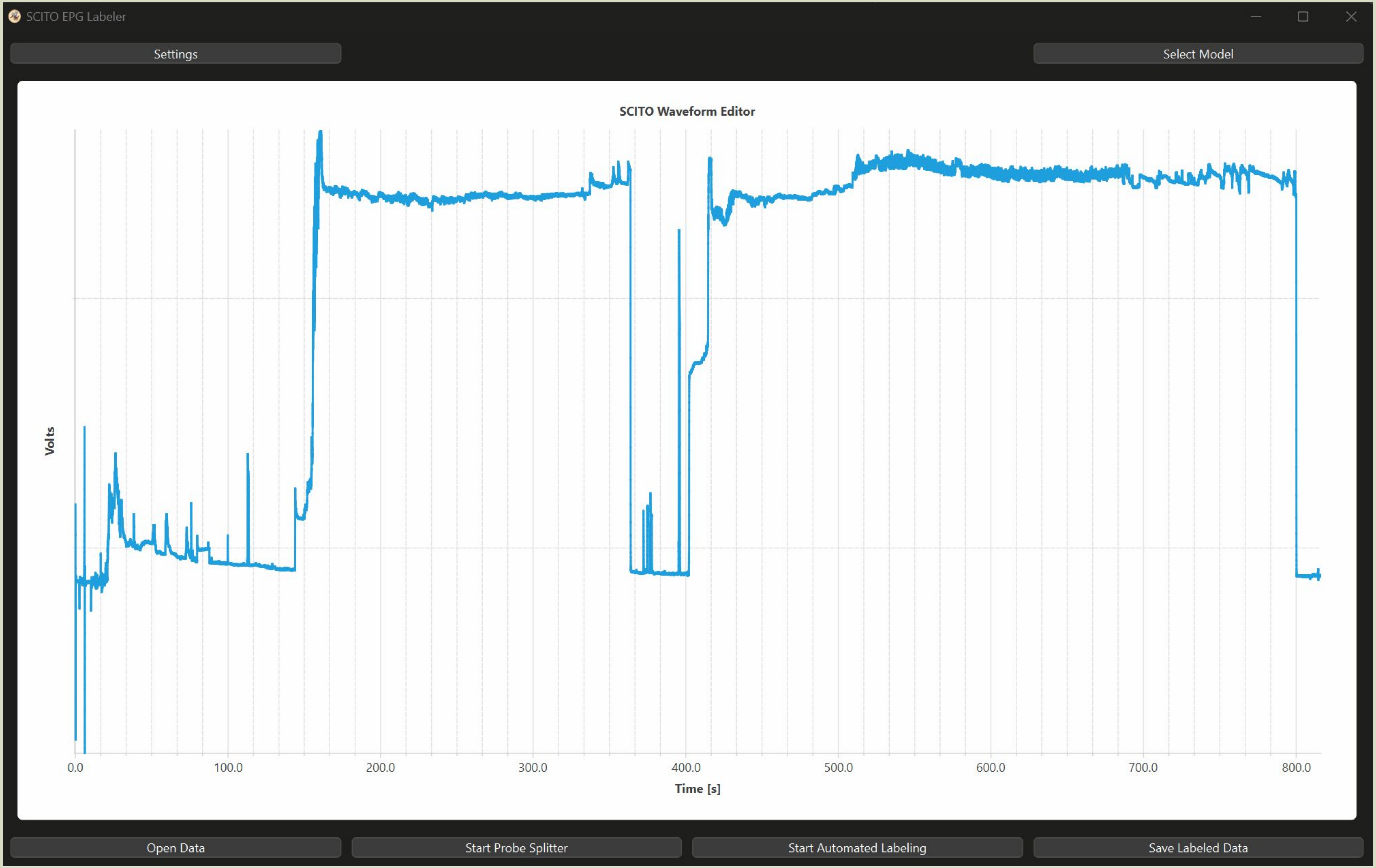


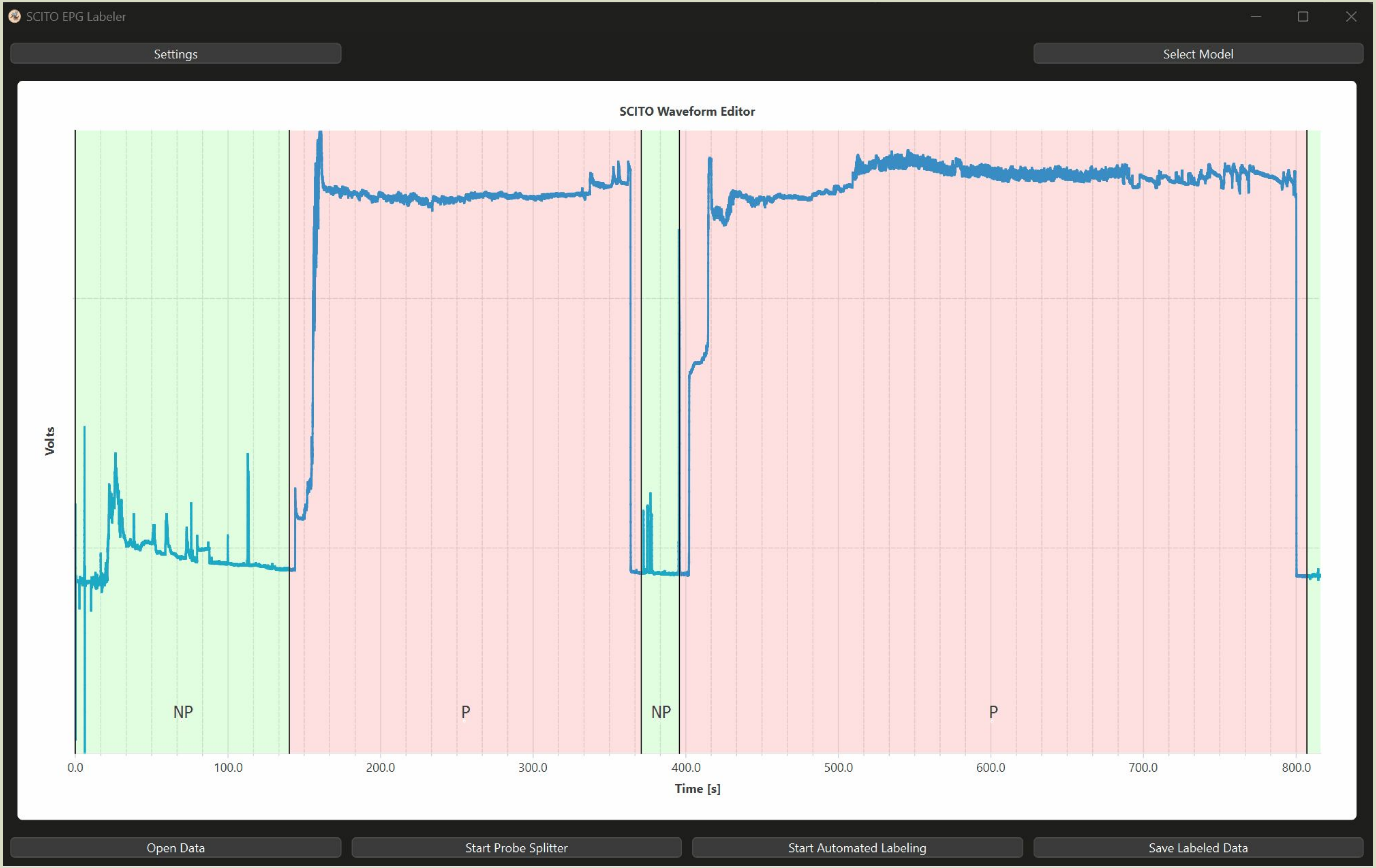
SCIDO

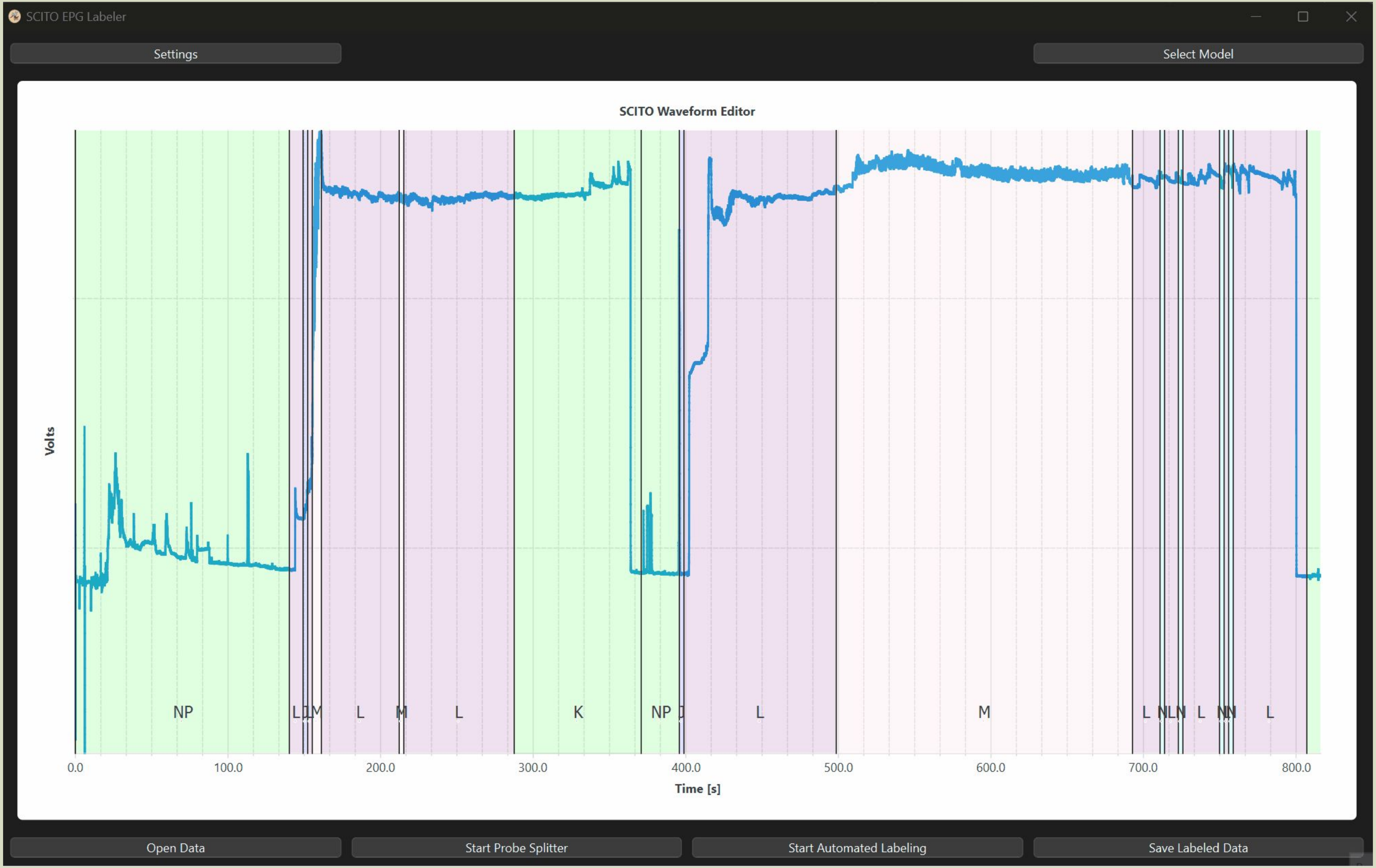

WINDAQ - CxT on hand 21Sep2021 no8.WDQ
File Edit View Search Scaling Transform XY Options Help
25.398
15.371
.151
1=1 Volt
pre1
post1
CxT pre1
CxT post1
CxT pre1
CxT post1
7.622
2=2 Volt
-14.602
-4.629
MAX 152.640 SEC(TBF) -878.630 SEC(TM) 14.8 %EOF T: 6.000 SEC/DIV

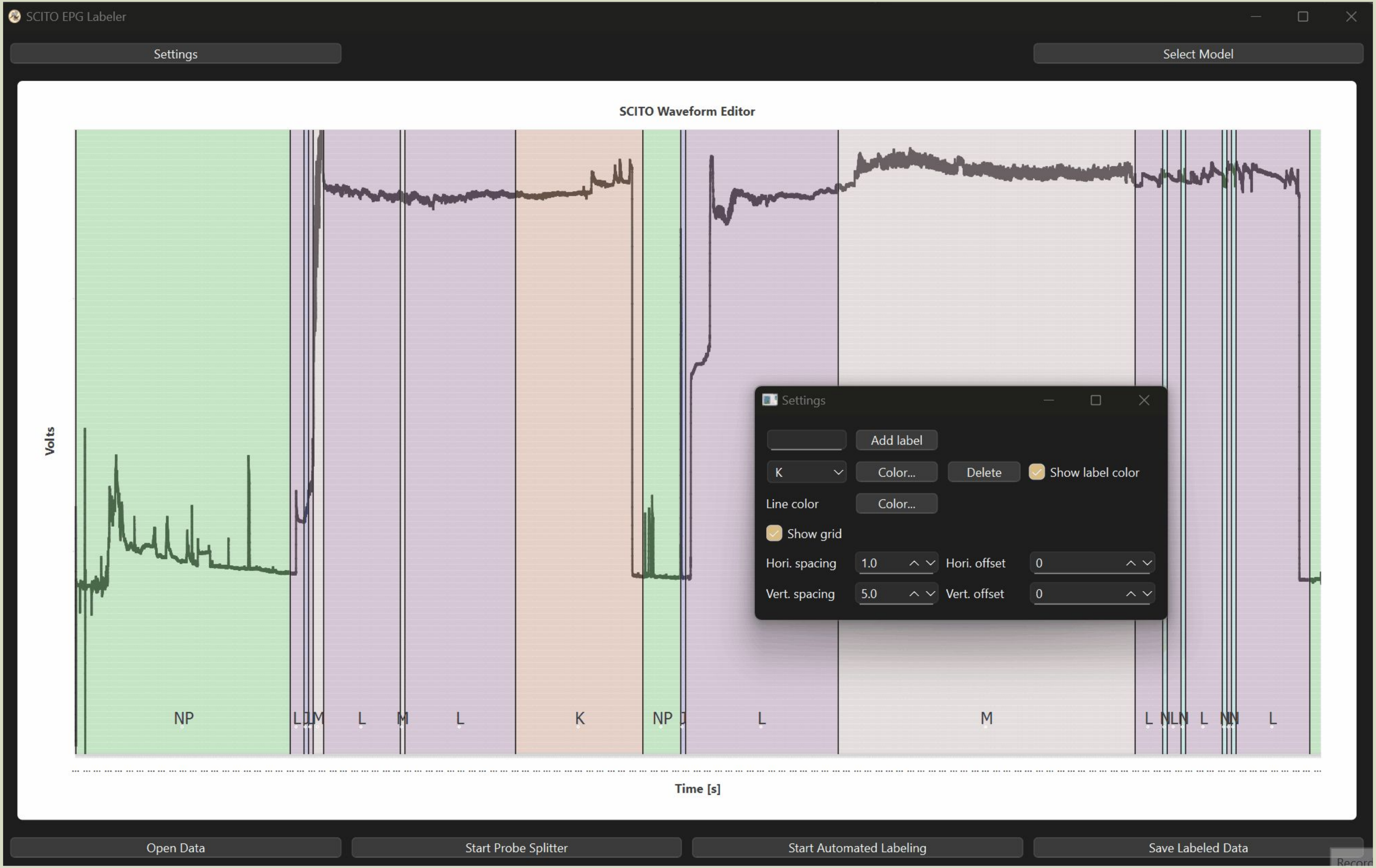


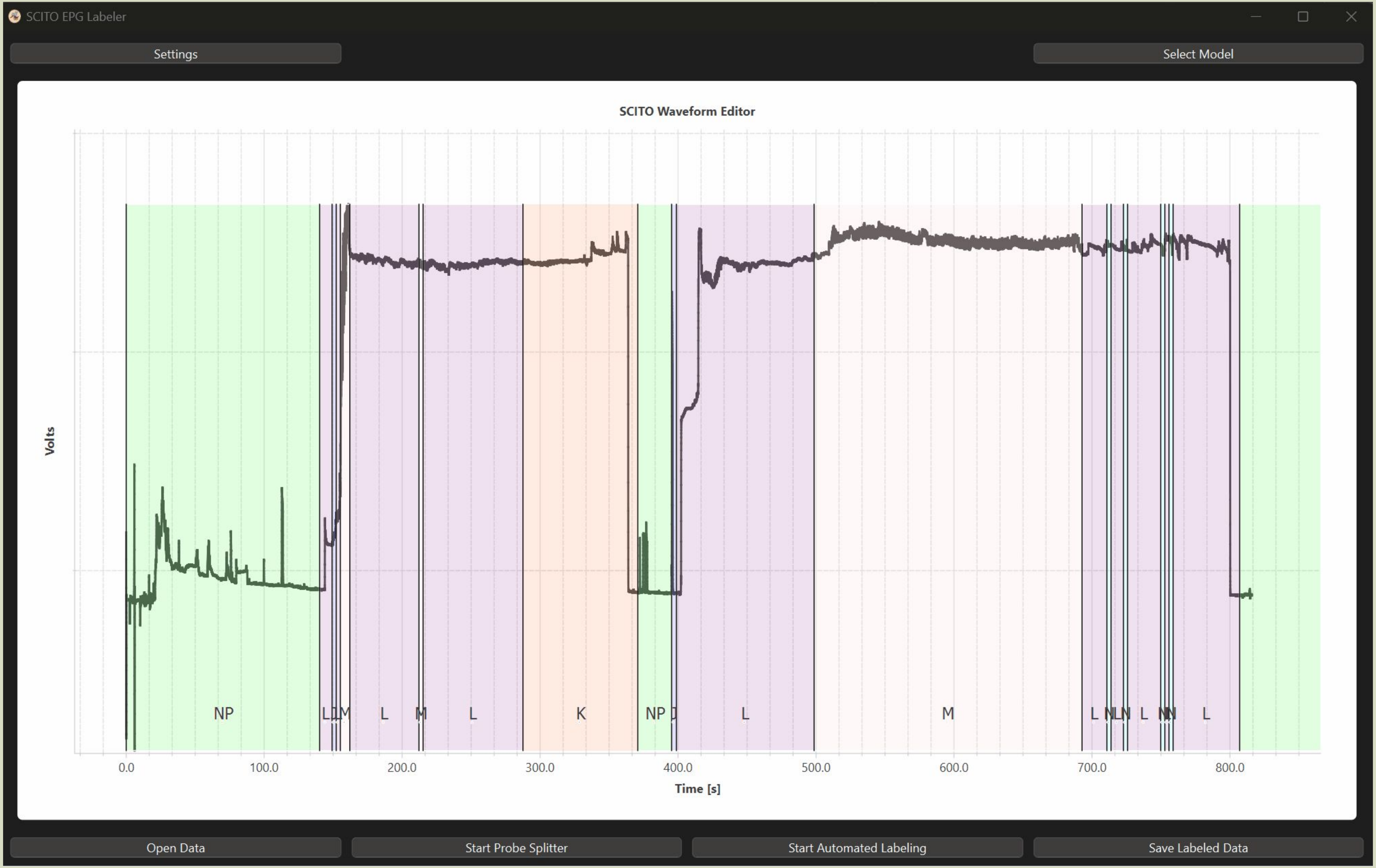
SCITO EPG Labeler
Settings
Select Model
SCITO Waveform Editor
Volts
0.0
NP
J
K
L
Z
NP
J
K
L
M
N
Z
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
Time [s]
Open Data
Start Probe Splitter
Start Automated Labeling
Save Labeled Data


SCITO EPG Labeler
Settings
Select Model
SCITO Waveform Editor
Volts
Time [s]
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
Open Data
Start Probe Splitter
Start Automated Labeling
Save Labeled Data


SCITO EPG Labeler
Settings
Select Model
SCITO 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


SCITO EPG Labeler
Settings
Select Model
SCITO Waveform Editor
Volts
NP
L
M
L
K
NP
L
M
L
L
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


SCITO EPG Labeler
Settings
Select Model
SCITO Waveform Editor
Volts
NP
L
M
K
J
Time [s]
Settings
Add label
K
Color...
Delete
Show label color
Line color
Color...
Show grid
Hori. spacing
1.0
Hori. offset
0
Vert. spacing
5.0
Vert. offset
0
Open Data
Start Probe Splitter
Start Automated Labeling
Save Labeled Data



SCITO EPG Labeler
Settings
Select Model
SCITO Waveform Editor
Volts
NP
L
M
L
M
L
K
NP
L
M
L
L
L
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

The Data

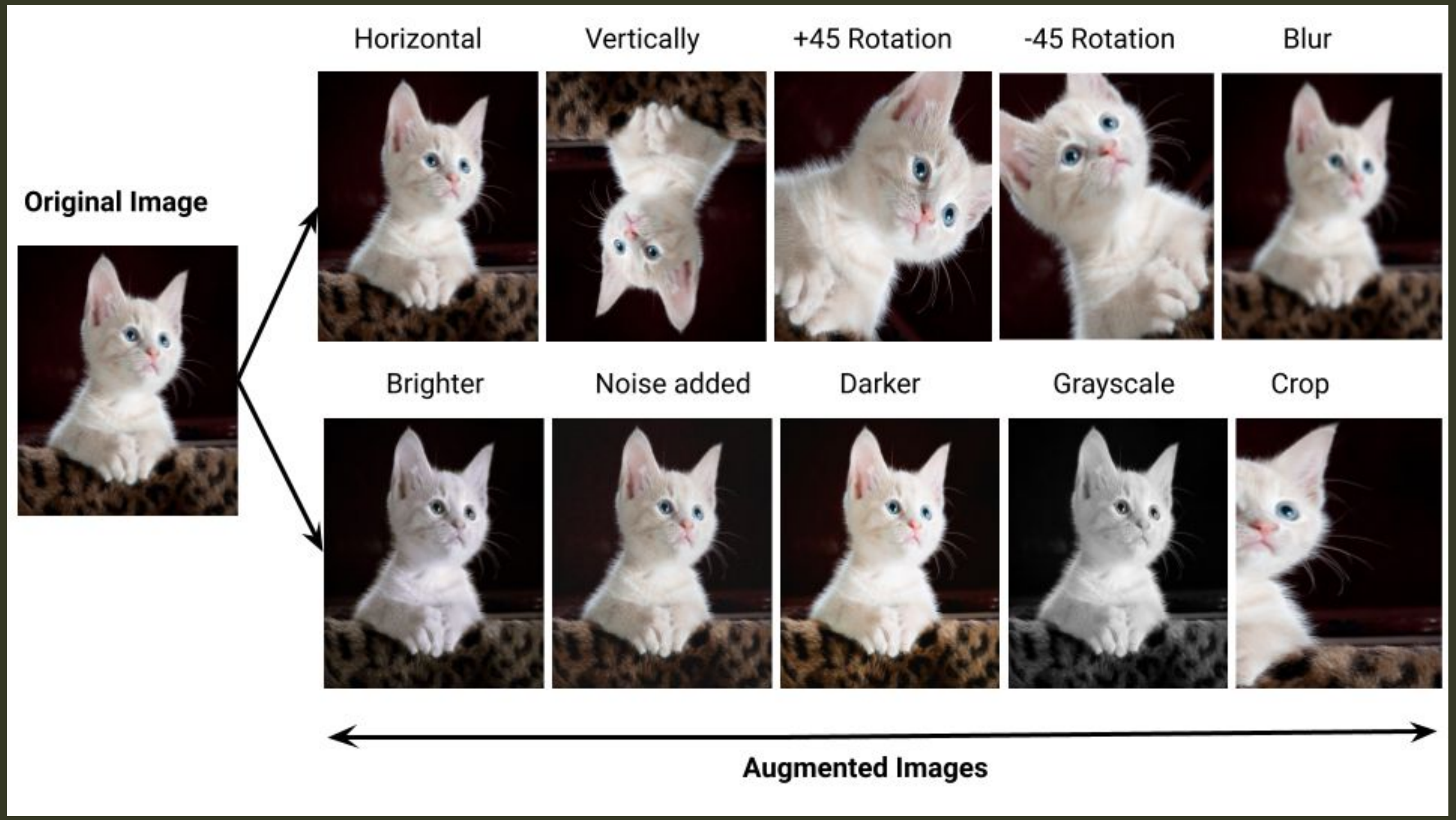
- 62 files
- 94 probes
- about 11 hours of data

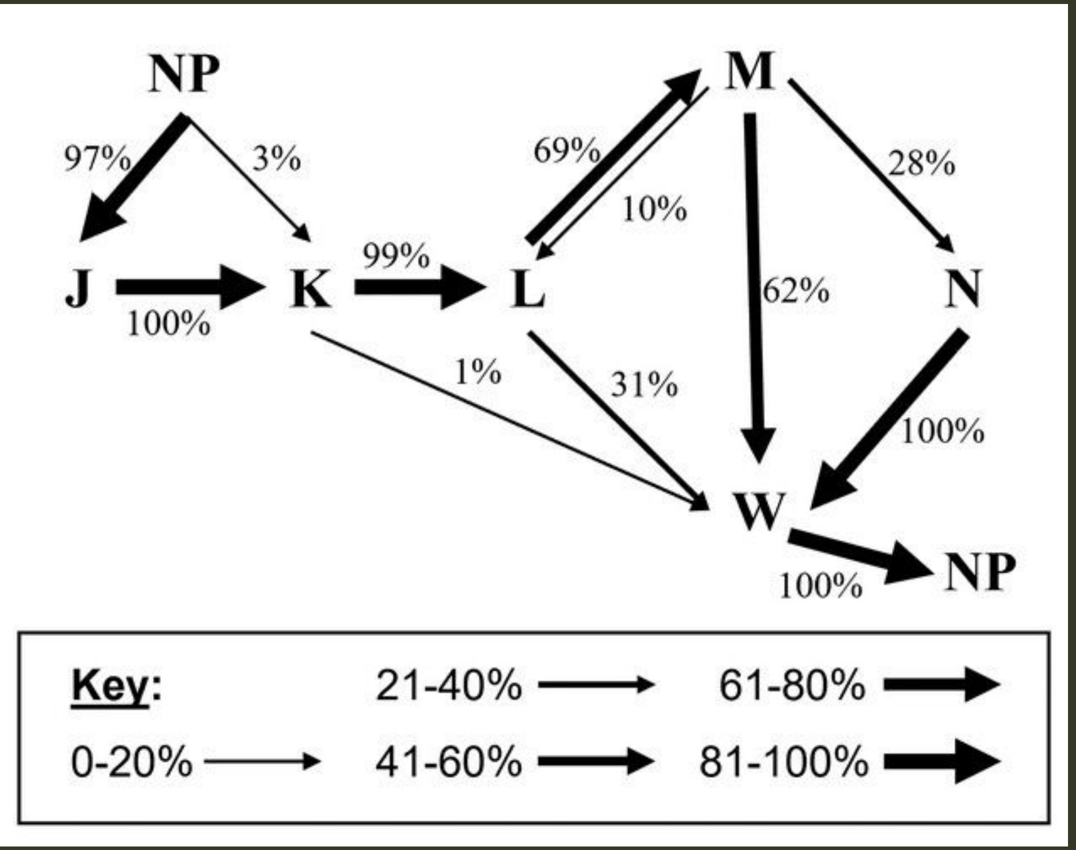
Imbalanced data

Feeding Stage: Hours of Data
'J': 0.2,
 'K': 0.1,
 'L': 4.8,
 'M': 5.1,
 'N': 0.7,
 'W': 0.01

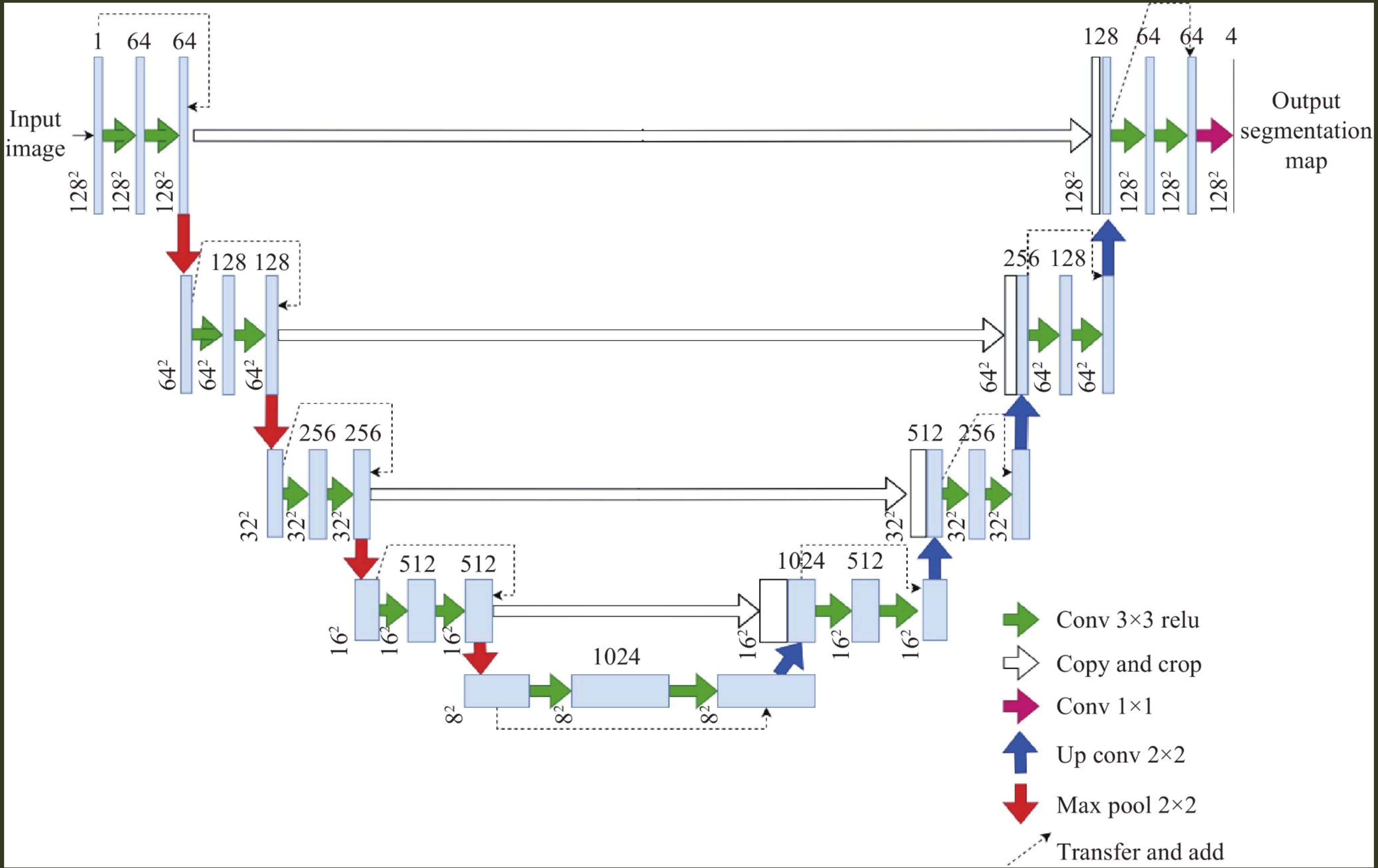
Data Augmentation

- Follow “rules” of the data generating process





The Main Model: UNet



The Main Model: UNet

Pros:

- Good at segmentation
- Large receptive field
- Highly expressive

Cons:

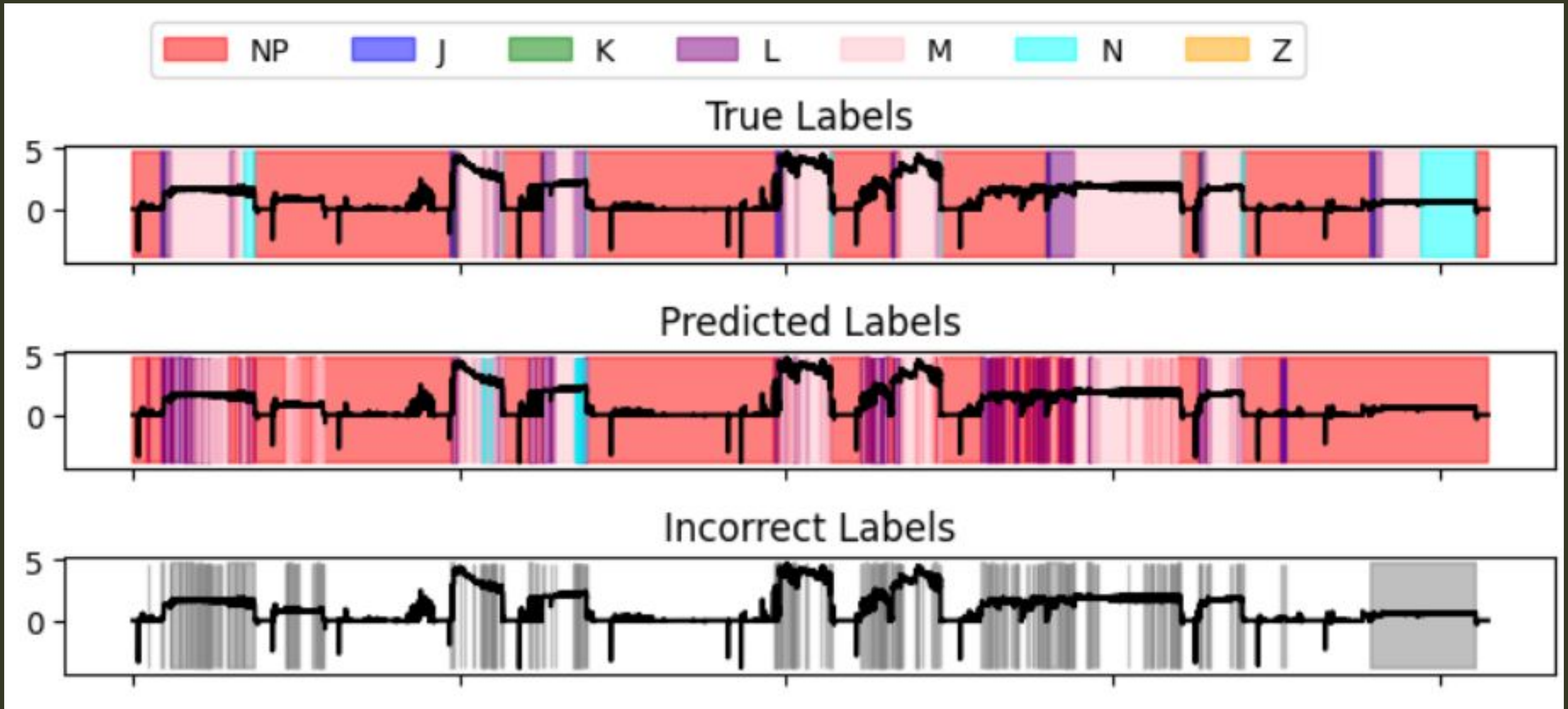
- Prone to overfitting

The Main Model: UNet

Accuracy: ~75%

state	precision	recall	fscore
J	0.77	0.94	0.84
K	0.17	0.28	0.21
L	0.65	0.98	0.78
M	0.98	0.79	0.88
N	0.01	0.00	0.00
W	0.00	0.00	0.00

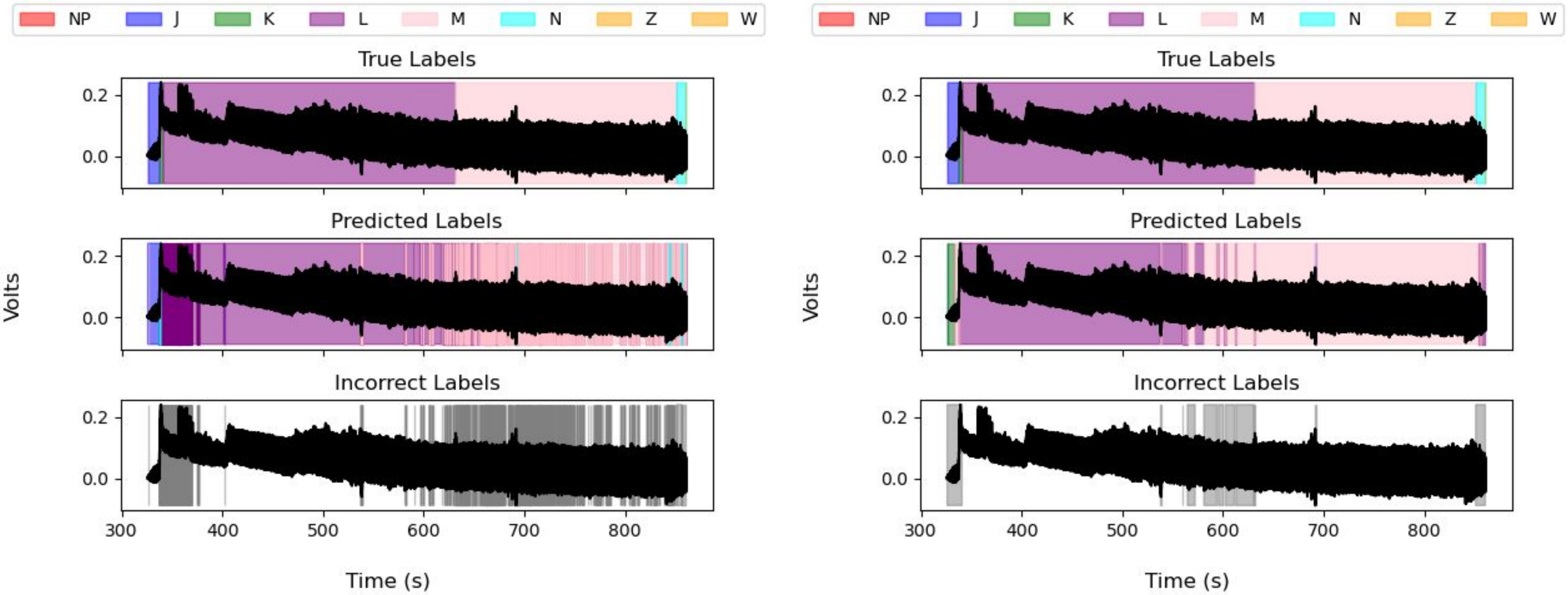
The Main Model: UNet



Post Processing

- Problem: “Barcodes”
- Solutions
  - Smoothing filter
  - Barcode cutter
  - HMM
  - HSMM

Example: HMM Postprocessor



Limitations

- Liaisons are experts in entomology but lack computer science experience
- Model output is never perfect, usually needs some manual adjustment
- Important to convey system limitations
  - Potential for negative impact on science if output is blindly trusted

Final steps

- GUI
  - Have list of features to implement from site visit feedback (mostly plot interaction)
- Machine Learning
  - Final optimizations (tuning hyperparameters)
  - Model descriptions and performance summary for use in liaison’s paper
- Refactoring to improve code cleanliness
- Documentation for users and developers (important for summer research students)

Thank you!

Questions/comments/suggestions?

We’d love to touch on ML models we’ve tried or next steps!

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