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

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

HARVEY MUDD COLLEGE

USDA

Raise your hand if…

Raise your hand if…

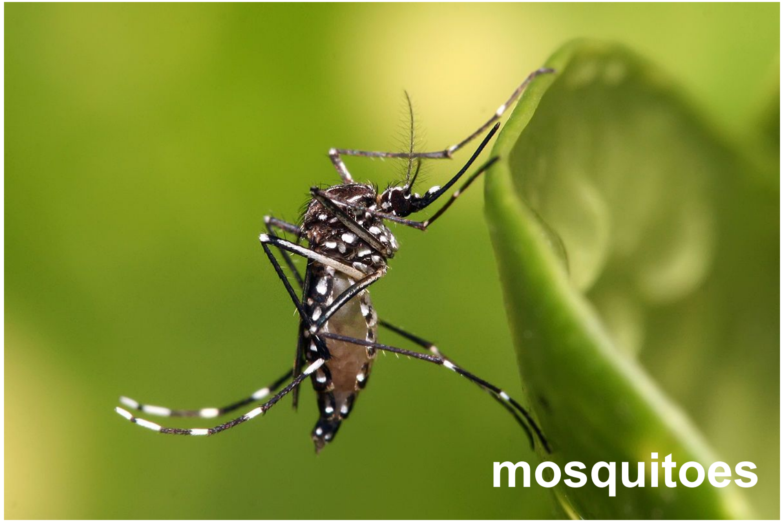
you’ve encountered a bug before

Let's talk about bugs!

Let’s talk about bugs!

Not this kind!

The real kind:



EPG

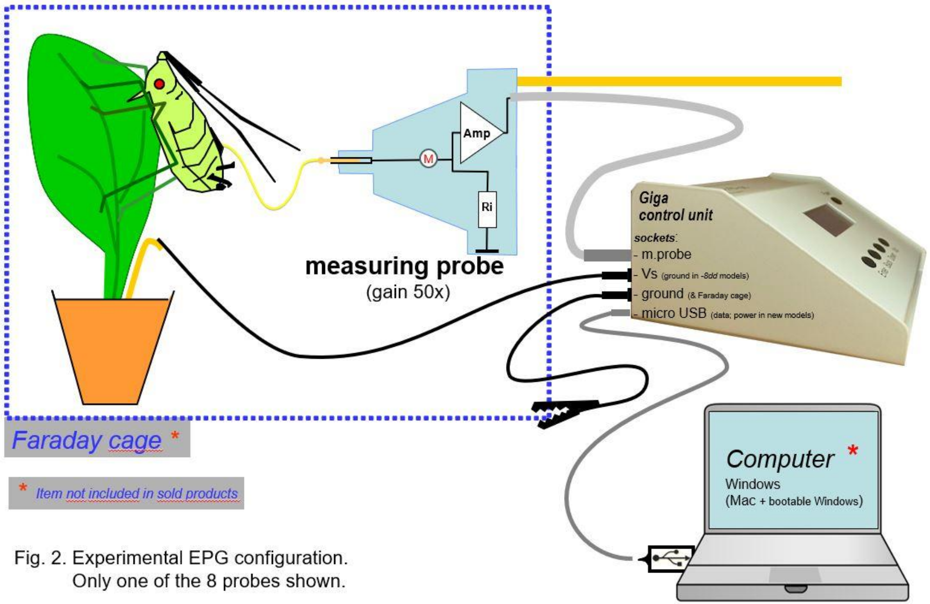


Fig. 2. Experimental EPG configuration.
Only one of the 8 probes shown.

EPG

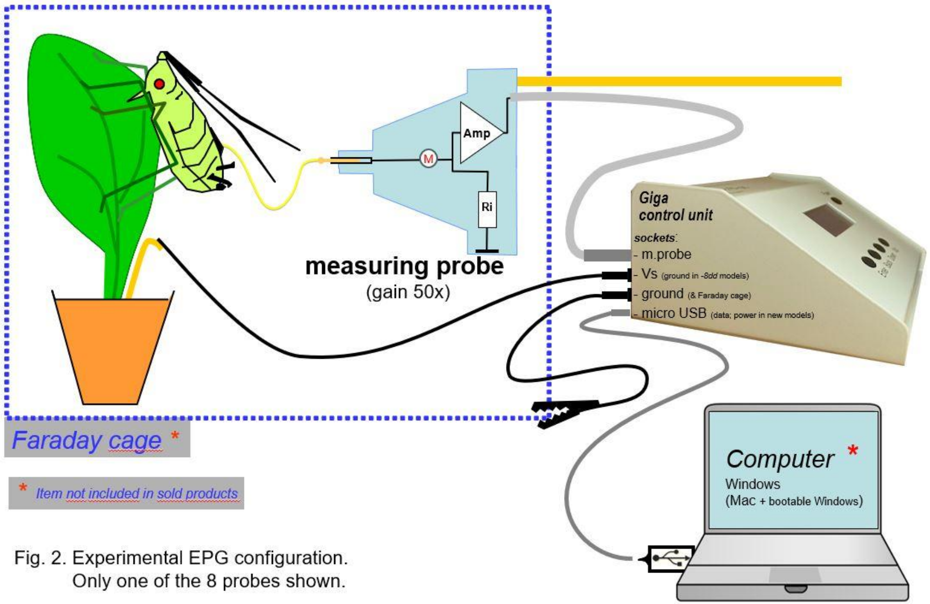
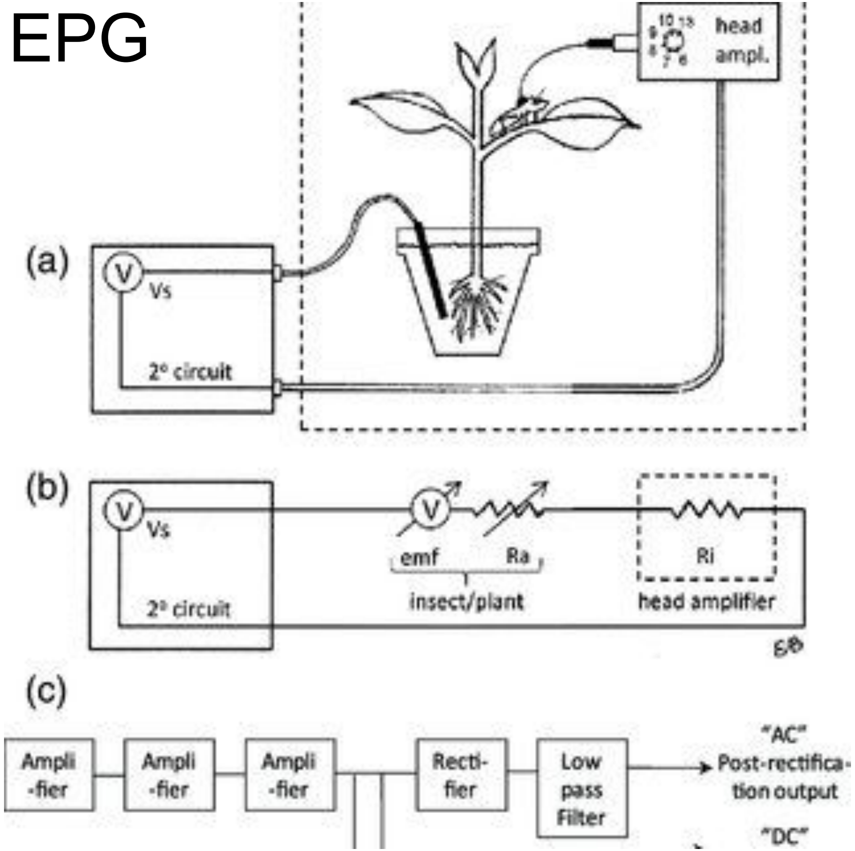


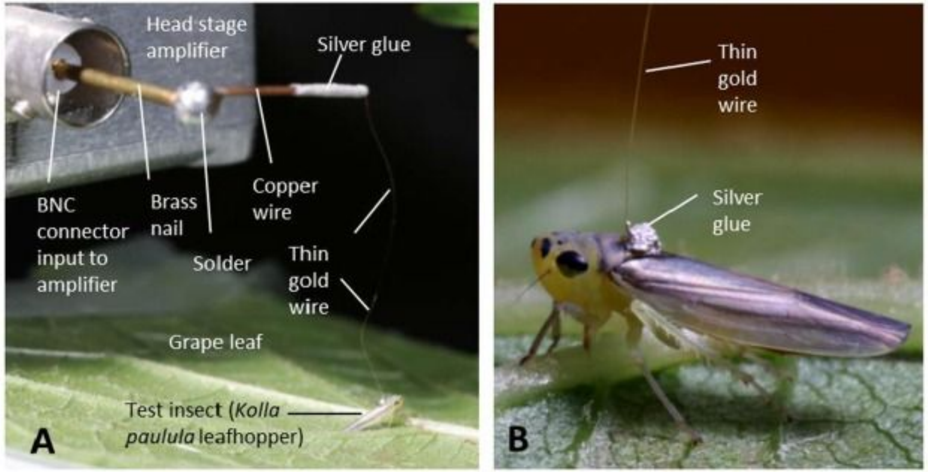
Fig. 2. Experimental EPG configuration.
Only one of the 8 probes shown.

Electrical Penetration Graph

EPG

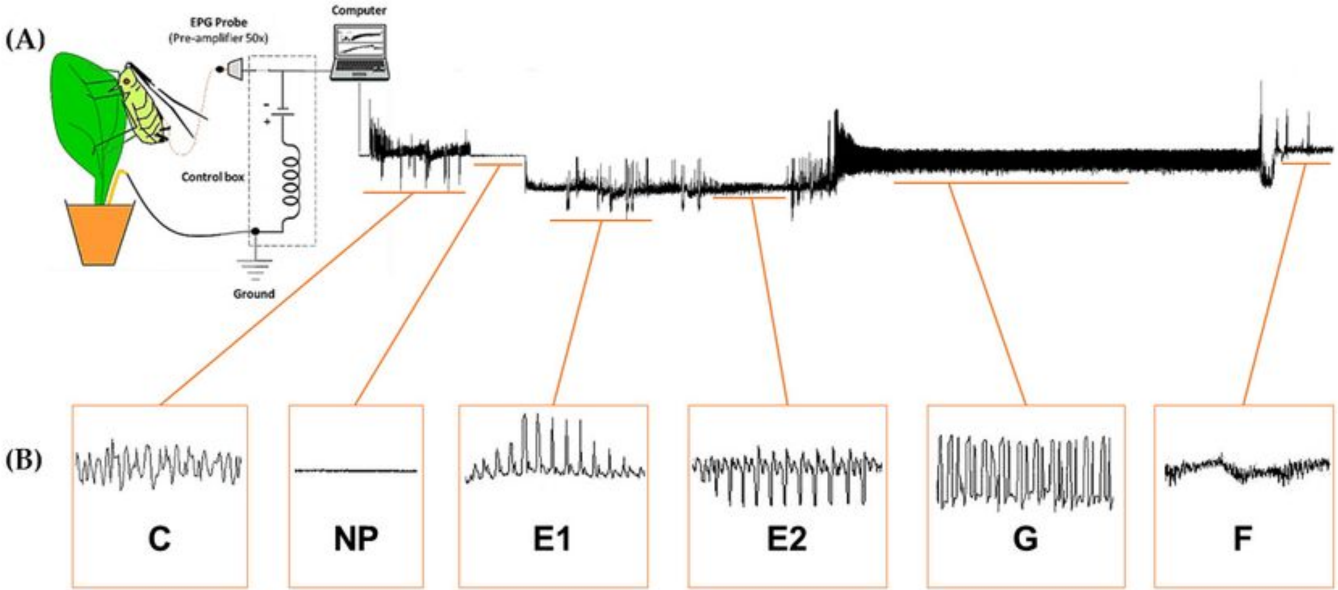


head
ampl.
(a)
Vs
2° circuit
(b)
Vs
2° circuit
emf
Ra
insect/plant
Ri
head amplifier
(c)
Ampli
-fier
Ampli
-fier
Ampli
-fier
Recti-
fier
Low
pass
Filter
"AC"
Post-rectifica-
tion output
"DC"



Head stage
amplifier
Silver glue
BNC
connector
input to
amplifier
Brass
nail
Solder
Copper
wire
Thin
gold
wire
Grape leaf
Test insect (*Kolla*
paulula leafhopper)
A
Thin
gold
wire
Silver
glue
B

EPG

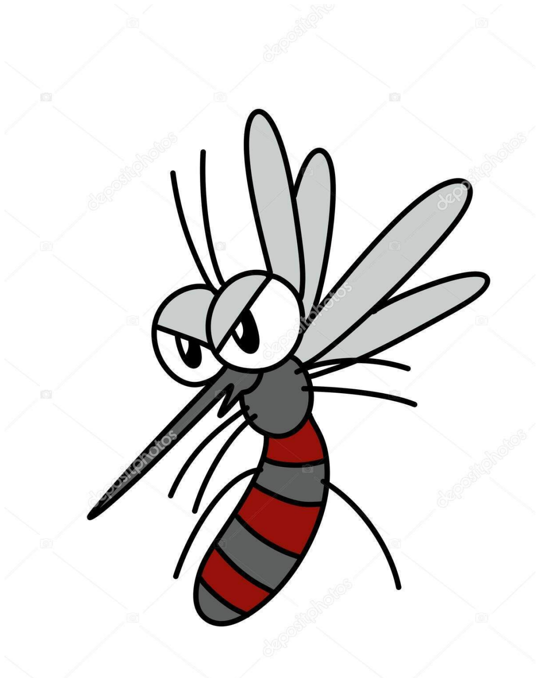

(A)
EPG Probe
(Pre-amplifier 50x)
Computer
Control box
Ground
(B)
C
NP
E1
E2
G
F

Who cares? (Here’s why you should!)

200,000 years

108 billion humans

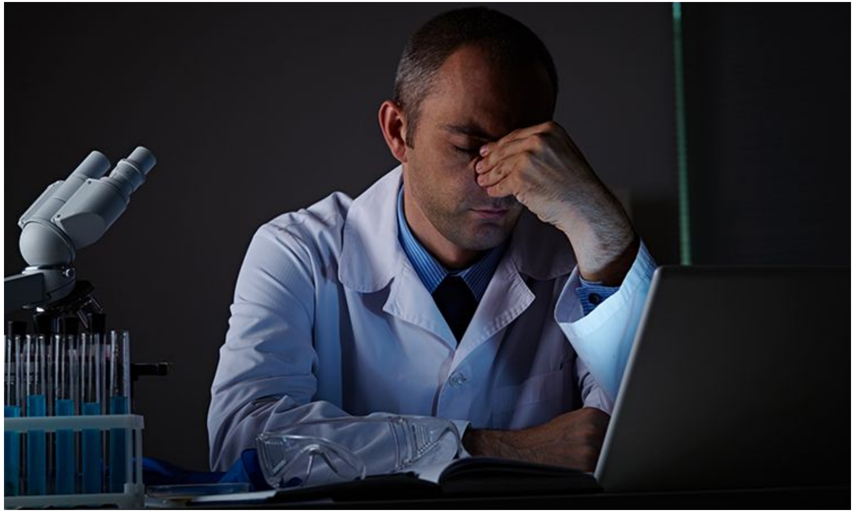
52 billion killed by mosquitos



[Our Deadliest Predator [source]]

How can we help?

- Waveform labelling is presently done manually
- This menial task takes ages, subject to inaccuracies
- Putting a bottleneck on research, delaying advancement



Problem Statement

We seek to automate electropenetrography annotation in order to accelerate entomological research on harmful pests.

Objectives

Define Performance Metrics

How do we measure our model's practical benefit to entomologists?

Develop the Model

Design and train a predictive model for waveform recognition

Develop an Interface

Researchers need to be able to upload data and receive labeled files

Stretch Goal

Build a tool that allows researchers to train new models for different species

Objectives

- Define a performance metric
  - Really hard to know how useful a given model will be to a researcher. While something simple like label accuracy is vaguely helpful, knowing where and why the model makes mistakes is key. We need to develop some basic metric (or set of metrics) to describe the performance of one of our models that has some alignment to how useful it will be to entomologists.
- Develop a machine-learning model that automatically labels data
  - As we’ve explain thus far, a big step is to actually design the predictive model that we will use to label the data automatically.
- Develop an interface for this model for use by entomologists
  - We want to design an easy-to-use interface for our audience, entomologists, who do not have a computer science background. They should easily be able to view the predictions of the model overlaid on the actual voltage data, potentially highlighting interesting areas such as low-confidence regions.
- Stretch goal
  - Additional tool which can re-train the model on either a new arthropod species or new data collected under different circumstances (i.e new machine)

Criterion for Success

It's tricky!

Previous models have hovered around an **80% accuracy rate**, which we have already achieved!

Accuracy isn't everything…

Success = producing a model that is *intuitively* usable by scientists

In terms of model performance, our project is based on **continuous improvement and balance of accuracy** between data regions

Criterion for Success

Accuracy isn’t everything…

Success = producing a model that is intuitively usable by scientists

In terms of model performance, our project is based on continuous improvement and balance of accuracy between data regions

- Knowing what success looks like in all these areas is tricky, since we can always have a better model or a more intuitive interface.
- By working closely with our sponsors, showing them prediction plots and summary statistics we can move towards a strong predictive model given the constrained data, compute, and time we are working with. We can iterate both the predictive model and interface based on what the entomologists themselves find most useful.
- Broadly, we think it will be a success if we have a predictive model that does reasonably well on all regions in the data and makes our sponsors happy, paired with an interface that our sponsors find intuitive and easy to use.

Context

- What is electropenetography
  - Picture of device
  - Picture of example waveform
  - What it is used for: learning about pests (mosquitos ew)

- Why our sponsors care about this project
    - Something about how waveforms have to be manually labeled and that takes a while and we’d like to be able to do that much faster since it puts a bottleneck on research

Context (Previous Work)

We aren't the first people to try to do this

Context (Previous Work)

We aren't the first people to try to do this

Waveforms Eavesdropping Prevention Framework: The Case of Classification of EPG Waveforms of Aphid Utilizing Wavelet Kernel Extreme Learning Machine

Yuqing Xing, Baofang Li, Lili Wu & Fengming Yan

To cite this article: Yuqing Xing, Baofang Li, Lili Wu & Fengming Yan (2023) Waveforms Eavesdropping Prevention Framework: The Case of Classification of EPG Waveforms of Aphid Utilizing Wavelet Kernel Extreme Learning Machine, Applied Artificial Intelligence, 37:1, 2214766, DOI: 10.1080/08839514.2023.2214766

Context (Previous Work)

We aren’t the first people to try to do this

Waveforms Eavesd... ...tion Framework:
The Case of Classi... ...rms of
Aphid Utilizing Wa...
Machine

Yuqing Xing, Baofang Li, ...

To cite this article: Yuqing Xing, Baofang Li, ...
Eavesdropping Prevention Framework: The Case of Classi...
Aphid Utilizing Wavelet Kernel Extreme Learning Machine, Applied Arti...
2214766, DOI: 10.1080/08839514.2023.2214766

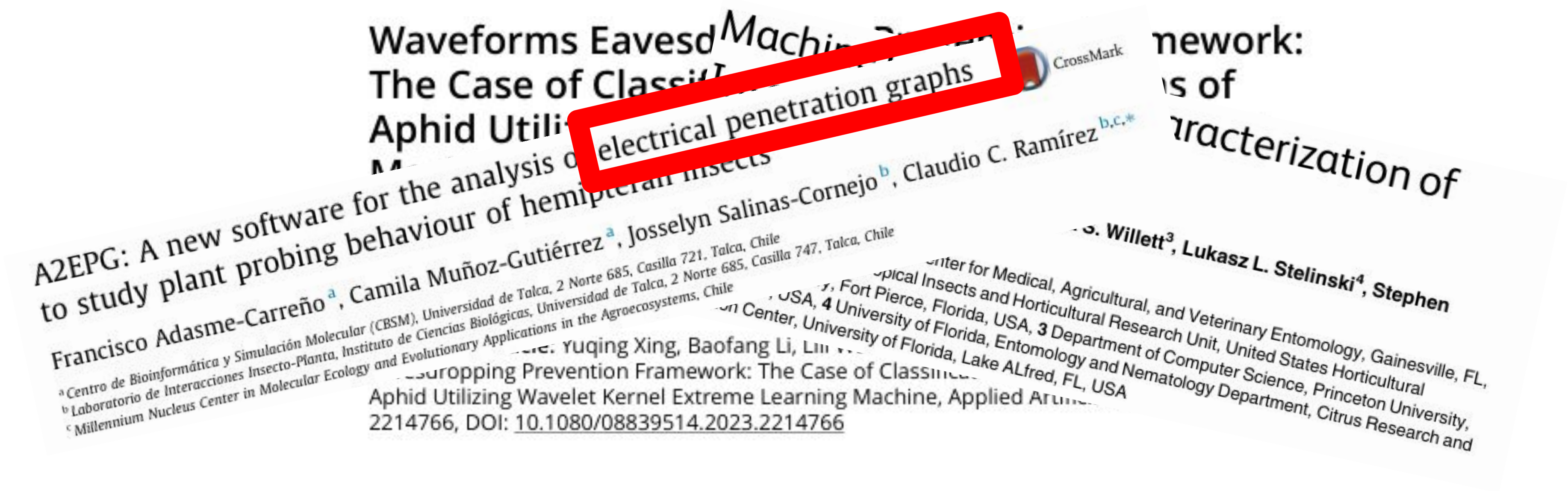
Machine Learning for Characterization of Insect Vector Feeding

Denis S. Willett[1*], Justin George[2], Nora S. Willett[3], Lukasz L. Stelinski[4], Stephen L. Lapointe[2]

1 USDA-ARS, Chemistry Unit, Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, FL, USA, **2** USDA-ARS, Subtropical Insects and Horticultural Research Unit, United States Horticultural Research Laboratory, Fort Pierce, Florida, USA, **3** Department of Computer Science, Princeton University, Princeton, NJ, USA, **4** University of Florida, Entomology and Nematology Department, Citrus Research and Education Center, University of Florida, Lake ALfred, FL, USA

Context (Previous Work)

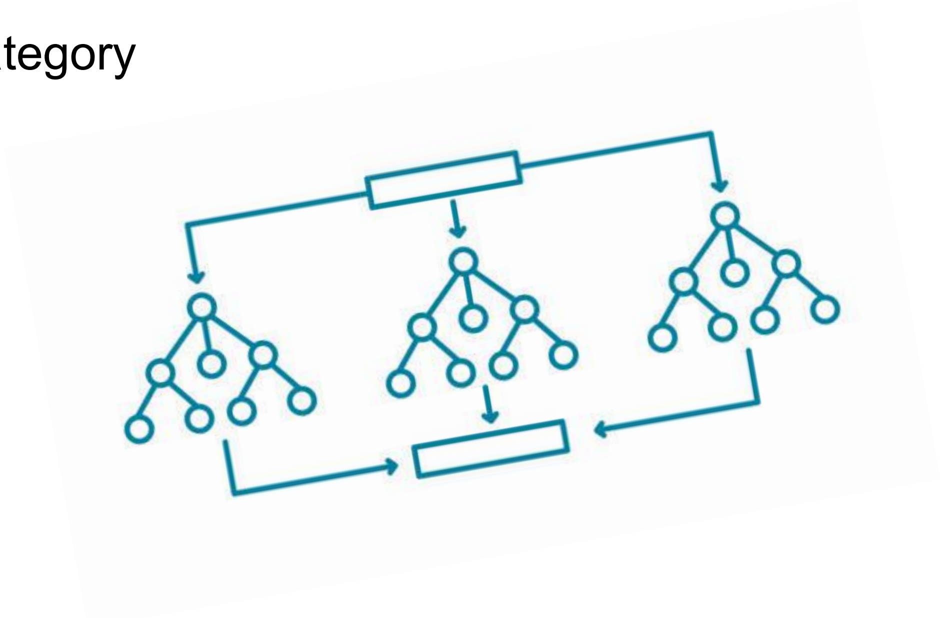
We aren't the first people to try to do this



Context (Previous Work)

How have people already attempted to solve the problem?

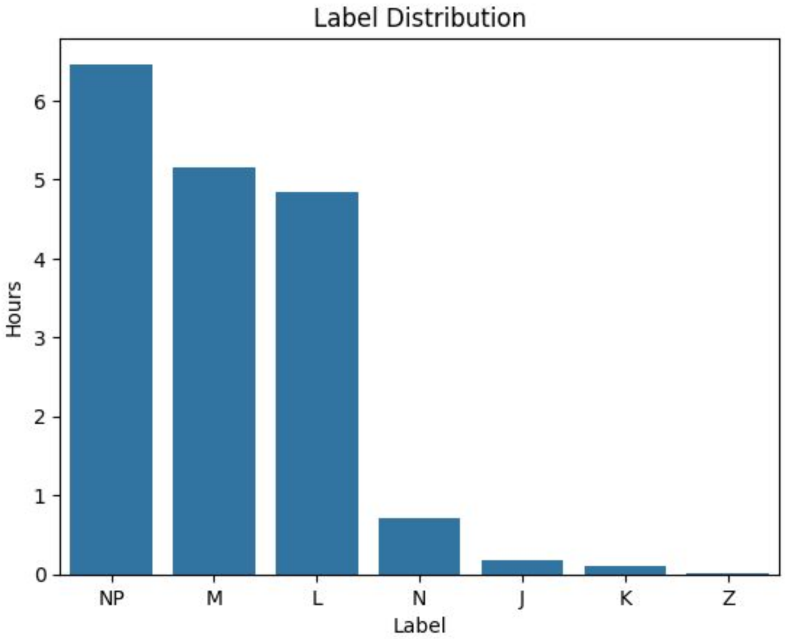
- Cut recording into chunks
- Calculate per-chunk statistics
- Use ***machine learning*** to predict category
    - Random forests



Context (Previous Work)

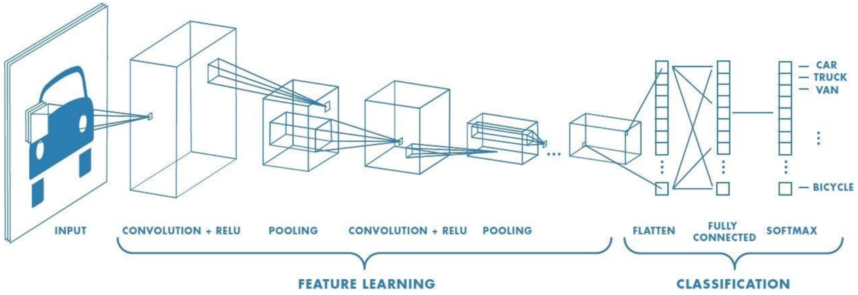
So what's wrong with those methods?

- Aphid-specific
- Fixed features
- Ignores context
- Poorly handles unbalanced data (like ours)



Context (Deep Learning)

- Local pattern recognition
- Well studied
- Efficient
- Good for automated feature extraction



Timeline

- November: Decision of machine learning approach
- December: Initial prototype of model and interface
    - Test prototype during site visit
- February: Improved prototype of model and interface
- April: Final prototype of model and interface

Plan

- October: Decision of Machine Learning Approach
- December: Initial Prototype of Model, Interface
- February: Second Prototype of Model, Interface
- April: Final Prototype of Model, Interface

Deliverables

- Machine learning model
  - Generate labels for a recording
  - Stretch goal: tool to retrain model for new data sets
- Interface
  - Allow user to upload file and view time-domain signal
  - Generate and overlay labels on the visualized signal using model

Ethics

- Sponsors have little computer science expertise
  - Must convey system limitations
  - If our software is sold, potential for scientific damage becomes greater

Acknowledgements

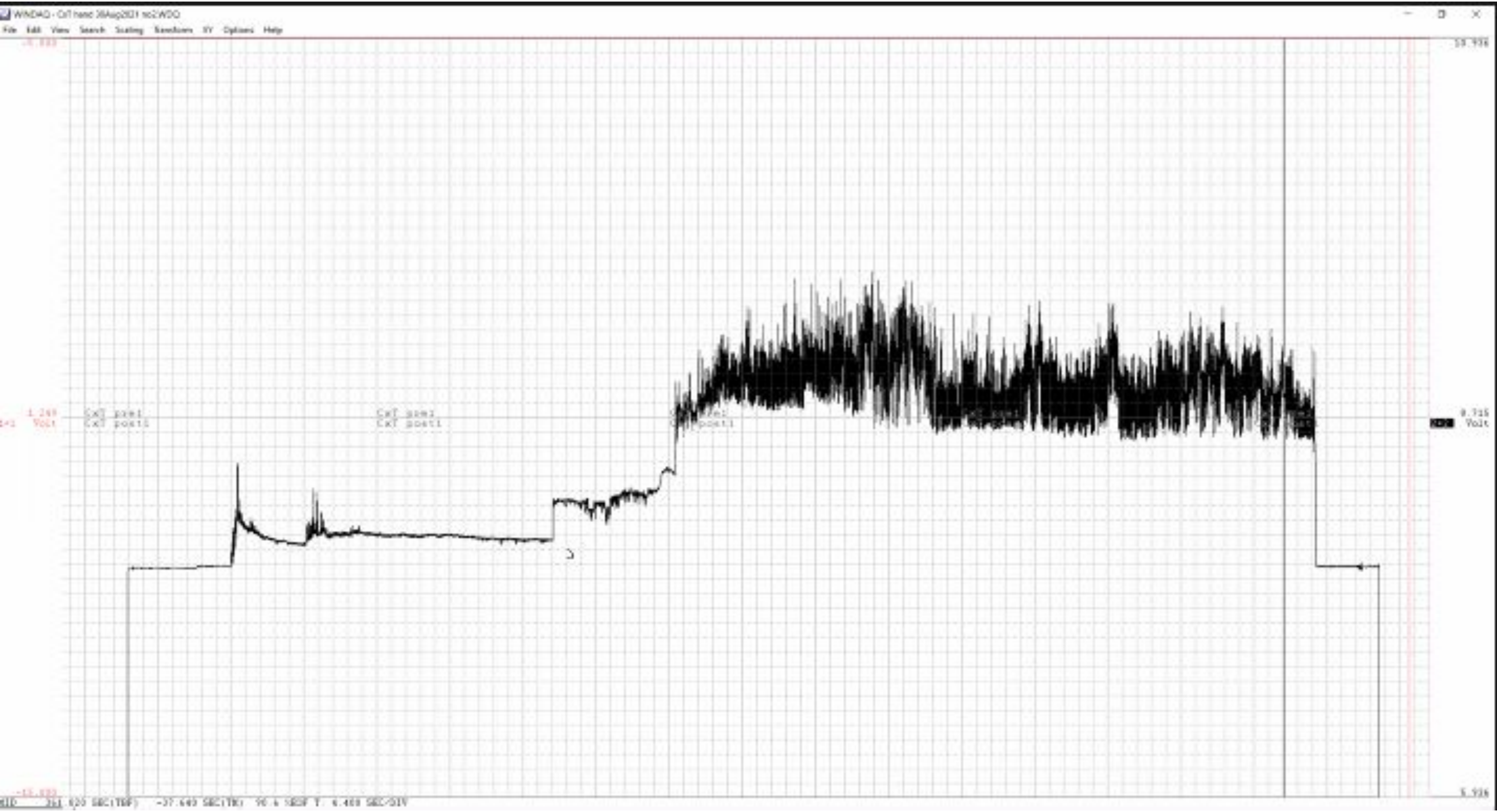
Funding
USDA (58-2034-3-445)
USDA (58-3022-4-034)
NSF (DBI - 2304787)

Thank You

Please clap now

Questions!

Please write down/share any suggestions for ML approaches that you might have!


File Edit View Search Scaling Transform XY Options Help
CxT post
Volt