



Self-sense: Semi-supervised Labeling of Sensor-Generated Time-Series Data



By: Tyler Lynn (tbl37@txstate.edu) & Tristan Pedro (jpp176@txstate.edu) Mentor: Dr. Vangelis Metsis

Department of Computer Science, Texas State University

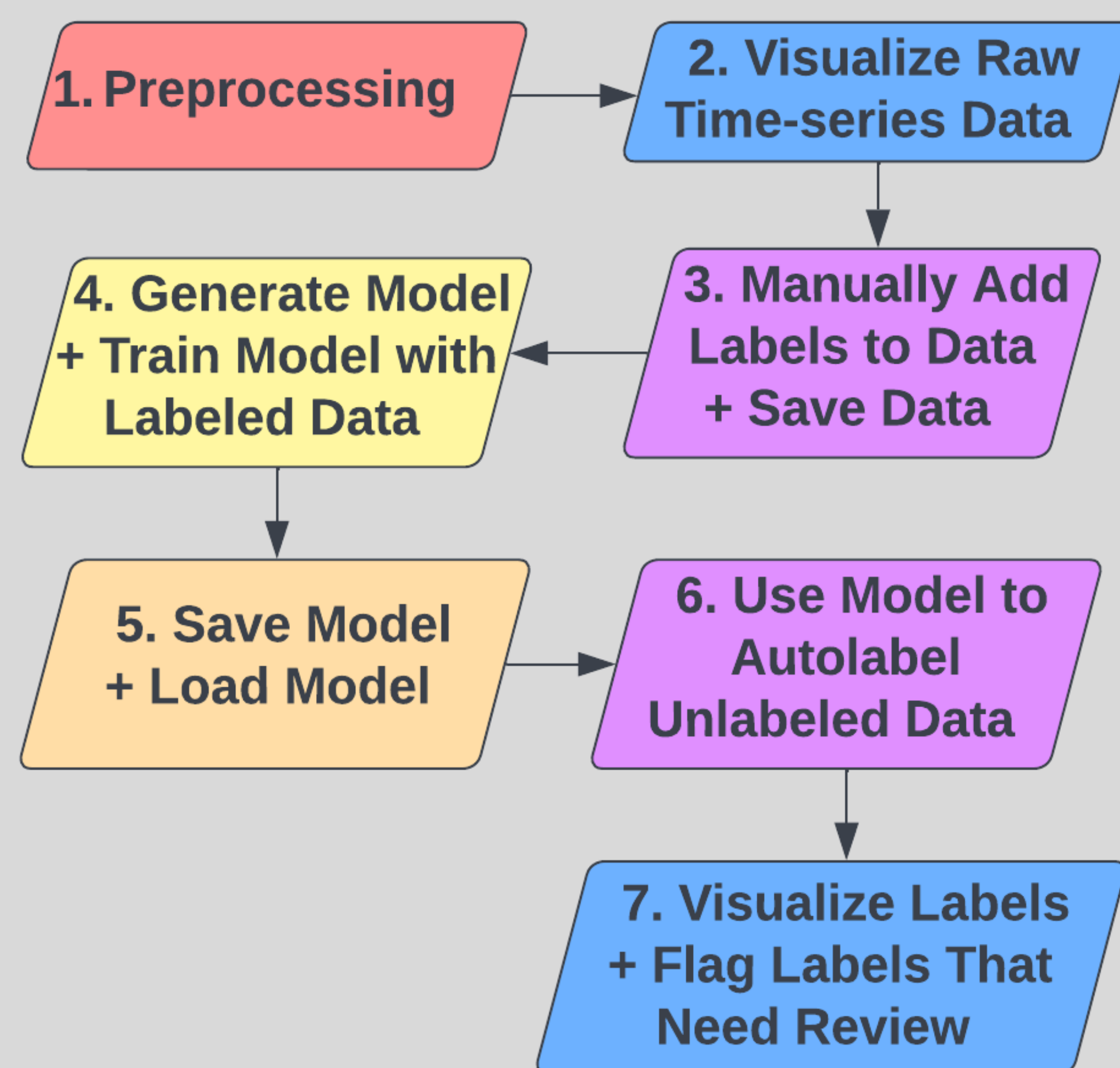
Introduction

Background: Time-series biomedical feedback data is data gathered from medical sensors that have a time dimension. E.g. EEG (electroencephalogram), EKG (electrocardiogram), Accelerometer, etc...

Motivation: Make it easier for biomedical researchers to visualize and create deep learning models from time-series bio-medical data.

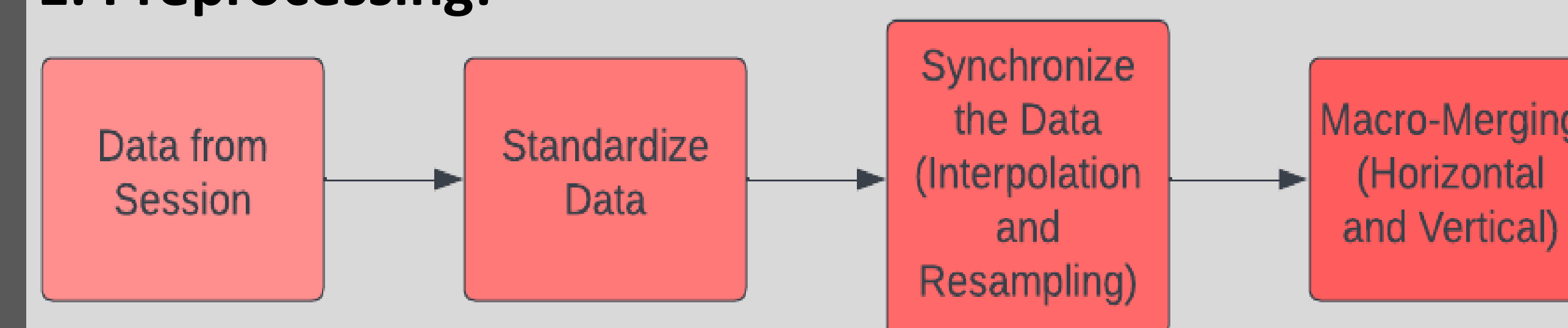
Objective: Create an application for visualizing, labeling, and creating models of biomedical data using deep learning.

Complete Project Pipeline



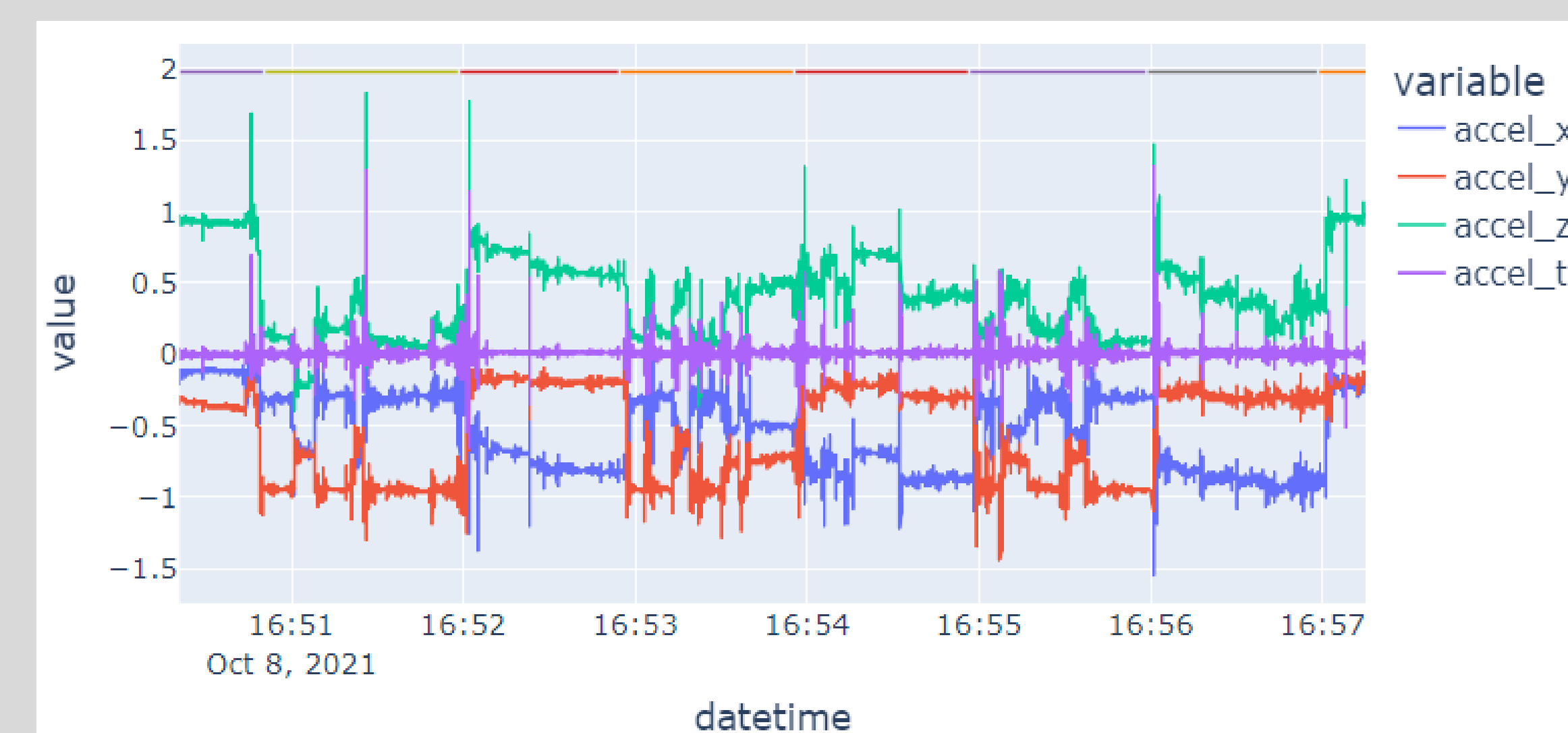
Functionality

1. Preprocessing:



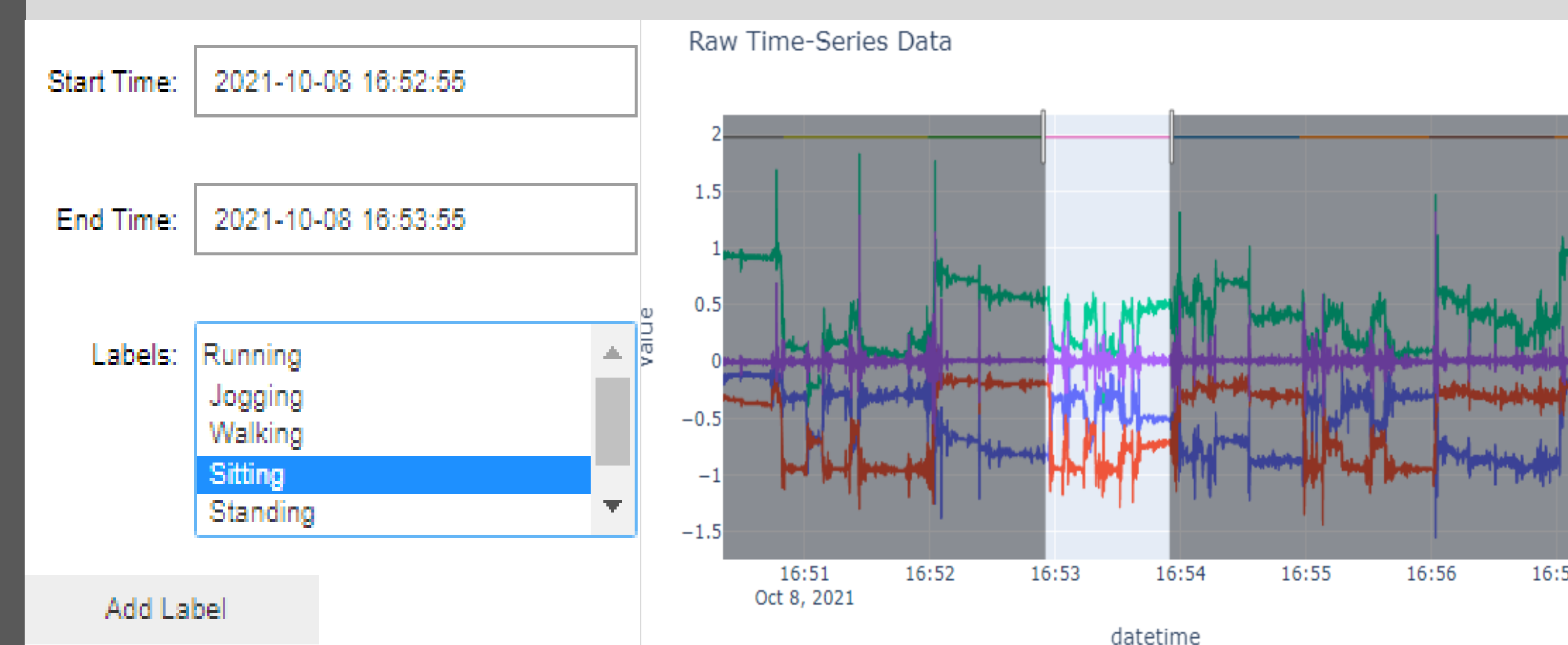
2. Visualize Raw Time-series Data:

Figure: Accelerometer Data



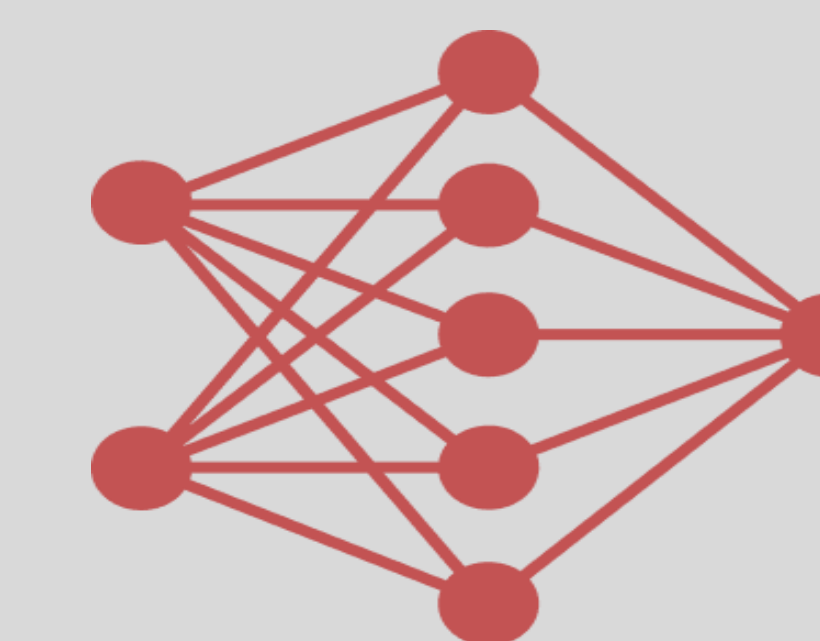
3. Manually Add Labels to Data & Save Data:

- Take user input from UI for start time, end time, and desired label. Apply that label for given time range



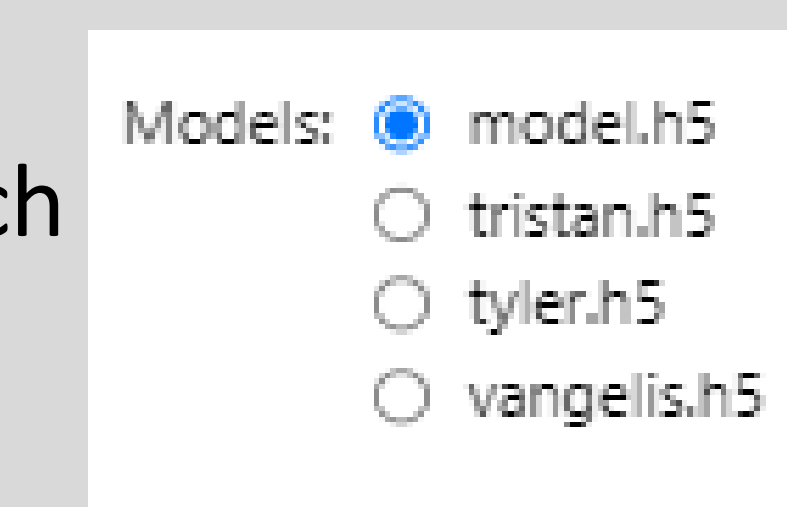
4. Train Model on Labeled Data & Save Model:

- Use specially created Numpy Array class to keep track of subject and time segment for each sample
- Train model on labeled data
- Name the model and save it



5. Load Model:

- Select which model to load and which data to label

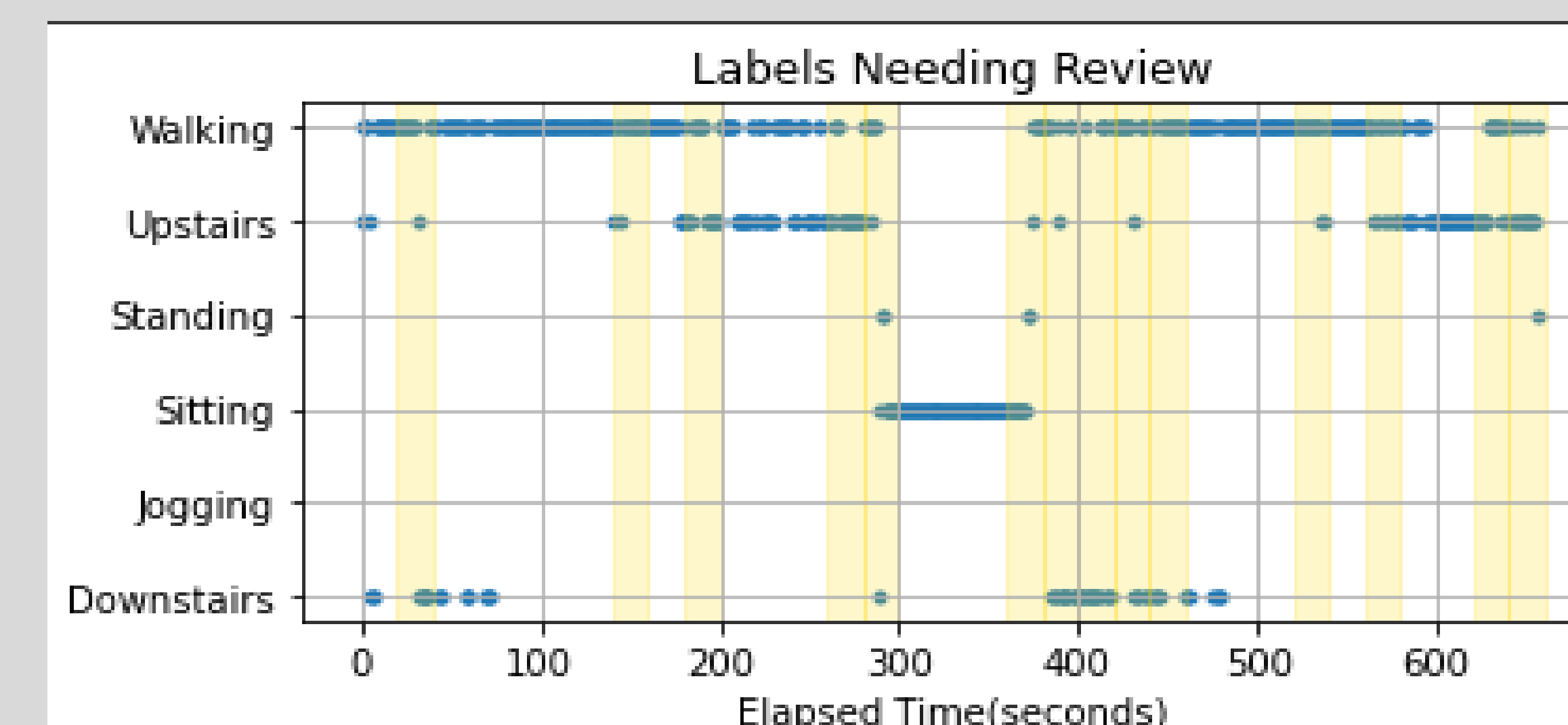


6. Upload & Label New Data:

- Use trained model to automatically label the new unlabeled data
- Use time from Numpy Array class to label each corresponding time segment on dataframe

7. Visualize Labels & Flag Labels That Need Review:

- Show the predicted activity labels in a scatter plot and highlight areas with yellow that need review



Conclusion

Results:

- Created a class for performing preprocessing on Empatica files
- Created application interface for project pipeline
- Used Jupyter Notebooks that are deployable through Voila (can be used to run, convert, and serve a Jupyter notebook as a standalone web app)
- First interface includes functionality for visualizing and labeling data, as well as creating and saving models
- Second interface includes functionality for uploading a model and using it to label new time series data

Future Work:

- Expand preprocessing library to more devices than Empatica (e.g. Bioradio, I-Phone Accelerometer)
- Expand Model library to include a more diverse Selection (e.g. LSTM, Transformers, RNN)
- Expand visualization suite (e.g. UMAP)
- Port project over to a desktop platform



Acknowledgements

NSF: This material is based upon work supported by the National Science Foundation under REU grants #2149950 and #1757893.

IMICS Lab: Special thanks to Lee Hinkle and Gentry Atkinson for their help.