Neuron Spike Identification with Machine Learning

Victoria Carlsten, Hao Chen, Claire Cropper, Shi Gu

**Abstract**—Epilepsy, a prevalent neurological disorder, impacts millions globally. Despite the array of treatments available, a significant number of patients remain non-responsive. This project involves the development of two machine learning algorithms that will accurately detect and identify the brain activity, or neuron spikes, that indicate active seizure activity. The first algorithm will focus on storing the identified seizure activity in order to retrospectively analyze after it has been collected. The second algorithm operates in real-time, allowing for the activity to be observed ongoingly, and for it to be actionable. The creation of two different app front-ends allows for the seizure activity to be visualized, and for any relevant parameters to be adjusted as necessary either as a method for analysis following the collection, or as it is actively occurring.

**Index Terms**—Clustering, classification, and association rules, Data and knowledge visualization, Machine Learning

—————————— ◆ ——————————

# 1 Need for This Project

OUGHLY 50 million people worldwide are diagnosed with Epilepsy, making it the most common neurological disease. Epilepsy results in frequent, recurring seizures due to unusual spikes of electrical activity in the brain. [1] In a healthy brain, nerve cells are able to communicate with each other through electrical activity. When a seizure occurs, the brain experiences an eruption of abnormal electrical signals that, for the time being, interrupts the usual electrical brain function. [2] 

While the development of anti-seizure medications has aided the majority of people in receiving relief from seizures, a third of the patients do not respond to current treatments. And, due to the risk of developing new neurological deficits, these patients are not viable candidates for surgery. [3] Drug-resistant epilepsy has resulted in newly developed brain stimulation methods as an alternative treatment option to decrease symptomatic seizures.

Several companies have contributed towards the research for development of a FDA-approved device that is able to deliver electrical stimulation to the brain in order to minimize the recurrence of seizures in patients resistant to medication. Deep Brain Stimulation (DBS) is a type of neuromodulation therapy that is surgically implanted and programmed to deliver electrical currents. By altering brain activity, seizures occur less frequently. However, the current open-loop neurostimulator delivers constant electrical stimulation in a predetermined cycle and lacks the capability to directly respond to seizures. [4] Implementing a contrasting closed-loop neuromodulation (CLN) system allows stimulation to be distributed when specific physiological states are encountered.

By leveraging machine learning techniques to automate retrospective and real time spike identification in electrophysiological readings, patients can receive neurostimulation when relevant signals are detected. Utilizing a closed-loop neuromodulation increases chances of achieving state-specific effects as well as minimizing side effects. Since stimulation only occurs when it is truly needed, CLN allows for a more efficient operation of the stimulus generator. [5]

This is the ultimate goal of these algorithms, as they allow for real-time, personalized control in treatment. Utilizing machine learning allows for each patient to have their own unique treatment plan based on their specific neural activity and seizure patterns. This is a significant step forward in the realm of epilepsy treatment. Personalized treatment ensures that patients receive the right amount of stimulation at the right times, minimizing potential side effects and maximizing therapeutic benefits.

——————

* *Victoria Carlsten CE’24. E-mail: carlsten@bu.edu*
* *Hao Chen CE’24. E-mail: ha0chen@bu.edu*
* *Claire Cropper CE’24. E-mail: ccropper@bu.edu*
* *Shi Gu EE’24. E-mail: bengushi@bu.edu*

# 2 Problem Statement, Methodology and Deliverables

## 2.1 Problem Statement

Our project is the development of two machine learning algorithms that accurately identify neuron spikes in electrophysiology data. Neuron spikes indicate neural activity and offer clues about resting state connectivity and stimulus response, offering insights into understanding brain function. One algorithm will work retrospectively for analysis after experiment. The second algorithm will work in real-time for Real-Time Spike Recordings (RTsr) for data collection validation. This project aims to leverage machine learning techniques to automate and enhance retrospective and real-time spike identification in complex electrophysiological recordings.

## 2.2 Methodology

1. Clustering Yang data,

2. clustering public data

3. Create preprocessing techniques to clean and transform raw public data into appropriate features for machine learning

4. Identify robust public datasets of labeled electrophysiology data containing spike and non-spike segments

5. Design, train and optimize retrospective machine learning models such that performance (F1, accuracy, precision, recall) is similar to or better than conventional spike labeling models

6. Evaluate and optimize machine learning models for data collected at Chen Yang Lab

7. Develop app front-end for visualizing detected spikes and changing parameters

8. Repeat 1-6 for RTsr (6 being optional for RTsr)

## 2.3 Deliverables

One python algorithm that:

1. Accepts raw electrophysiology data, and outputs labeled spike-trains and neuron clusters with confidence of labeling
2. App front-end for visualizing spikes and altering parameters (such as clustering conductivity thresholds, accepted sensitivity, etc.)

​A second python algorithm that:

1. Accepts RTsr and labels spike-trains + neuron clusters in real time
2. ​App front-end for real-time electrophysiology model (optional)

# 3 Visualization

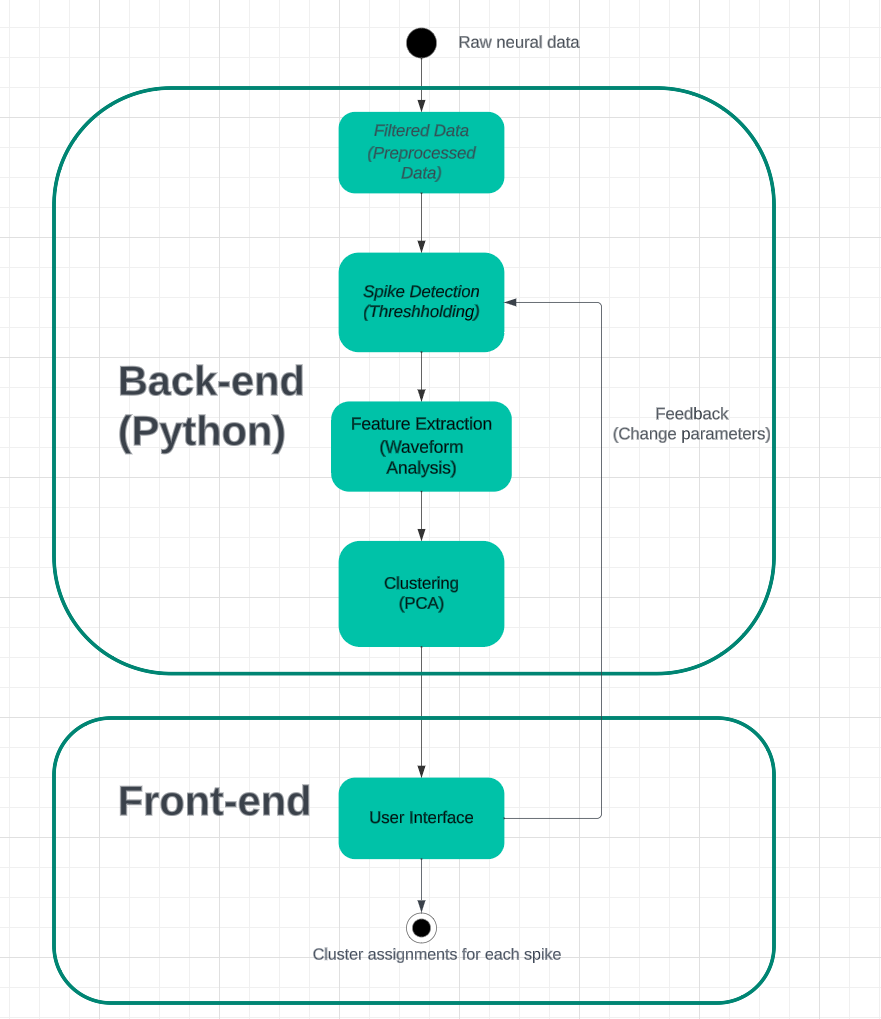


Fig. 1 Block diagram of the application

Our application is structured into two primary segments: the backend and the frontend. The backend, written in Python, accepts and processes raw electrophysiology data from detection devices. It filters noise, detects spikes using thresholding, extracts pivotal features, and employs machine learning techniques, like PCA, for neuron cluster classification. This refined data is then channeled to the frontend, which visualizes labeled spike-trains and neuron clusters with confidence of labeling. Additionally, an interactive user interface in the frontend lets users tweak processing parameters, offering tailored visualization results. The programming language for the frontend is based on the platform, which is still under consideration.

# 4 Competing Technologies

Scientists and industries across several research groups are developing spike sorting algorithms for the purpose of efficient grouping of spikes into clusters based on the similarity of their shapes . The most significant among these are the NTM method, SpikeDeeptector, and the Offline Sorter by Plexon.

## 4.1 Normalized-template-matching (NTM) method

Scientists in University of California Berkeley Helen Wills Neuroscience Institute have developed a simple, data-driven spike detection method using a scaled form of template matching, based on the sliding cosine similarity between the extracellular voltage signal and mean spike waveforms of candidate single units. Spike-sorting with the NTM method is a two-part process. First, an initial round of spike detection is performed using a standard fixed voltage threshold, and spike sorting is performed to cluster spike waveforms by their shape. This is followed by a second round where spike detection is performed via the NTM template matching method, and the goal is to maximize spike detection for these specific candidate single units. Detected spikes are then clustered by spike sorting, as in the standard method. They use one particular spike clustering method here, but NTM can be coupled with any clustering method. [6]

## 4.2 SpikeDeeptector

Scientists led by Dr. Muhammad Saif-ur-Rehman proposed a novel algorithm based on a new way of feature vector extraction and a deep learning method, which they called SpikeDeeptector. SpikeDeeptector considers a batch of waveforms to construct a single feature vector and enables contextual learning. The feature vectors are then fed to a deep learning method, which learns contextualized, temporal and spatial patterns, and classifies them as channels containing neural spike data or only noise. They used neural networks as the machine learning algorithm, and found out CNN (Convolutional Neural Networks) performs a more accurate result. The cumulative evaluation accuracy was 97.20% on 1.56 million hand labeled test inputs. Significance. The results demonstrate that SpikeDeeptector generalizes not only to the new data, but also to different brain areas, subjects, and electrode types not used for training. The result came out with a high accuracy with large amounts of samples. However, since they used neural networks for deep learning, this method can be costly and time-consuming. [7]

## 4.3 Offline Sorter by Plexon

Offline Sorter by Plexon is a software for viewing and classifying action potential waveforms (spikes) previously collected from single electrodes, stereotrodes and tetrodes. Spikes are displayed as points in either two-dimensional or three-dimensional feature space, where a variety of manual, semi-automated or fully-automated clustering techniques can be applied to classify the spikes. The different mode of clustering uses algorithms such as K-Means and E-M (Expectation Maximization). The export file type can be MATLAB, Excel, text file or NeuroExplorer. One advantage of it can be that it has a user interface which is user-friendly. We also have it and we are trying to develop software using Python and different algorithms (e.g. PCA) from theirs. [8]

# 5 Engineering Requirements

## 5.1 Data Collection and Identification

1. The first algorithm will collect and store electrophysiological data as it is being recorded, and categorize and label spike-trains and neuron clusters for analysis post collection.
2. The second algorithm will collect and store electrophysiological data as it is being recorded, and categorize and label spike-trains and neuron clusters in real time. As real time spike recordings (RTsr) are collected, it must be actionable data.

## 5.2 Visualization

1. The first app front-end will indicate (1) detected spikes that are identified from data collected, and (2) any altering parameters, such as clustering conductivity thresholds, accepted sensitivity, or additional parameters from post-collection.
2. The second app front-end will indicate (1) detected spikes that are identified from data collected, and (2) any altering parameters, such as clustering conductivity thresholds, accepted sensitivity, or additional parameters in real time.

# 6 Appendix A References

1. “Epilepsy,” World Health Organization,

https://www.who.int/news-room/fact-sheets/detail/epilepsy (accessed Oct. 20, 2023).

1. “Epilepsy,” Mayo Clinic,

https://www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093 (accessed Oct. 20, 2023).

1. “Brain stimulation therapies for epilepsy,” National Institute of Neurological Disorders and Stroke,

https://www.ninds.nih.gov/about-ninds/impact/ninds-contributions-approved-therapies/brain-stimulation-therapies-epilepsy (accessed Oct. 20, 2023).

1. A. B. K. MD, A. By:, E. K. MD, R. By:, and Adriana Bermeo-Ovalle MDStephan U. Schuele MD, “Deep Brain stimulation,” Epilepsy Foundation,

https://www.epilepsy.com/treatment/devices/deep-brain-stimulation (accessed Oct. 20, 2023).

1. S. Zanos, “Closed-loop neuromodulation in physiological and Translational Research,” Cold Spring Harbor perspectives in medicine, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6824403/#:~:text=In%20a%20closed%2Dloop%20neuromodulation,real%20time%20(adaptive%20neurostimulation) (accessed Oct. 20, 2023).
2. K. J. Laboy-Juárez, S. Ahn, and D. E. Feldman, “A normalized template matching method for improving spike detection in extracellular voltage recordings,” Nature News, https://www.nature.com/articles/s41598-019-48456-y (accessed Oct. 20, 2023).
3. M. Saif-ur-Rehman1 et al., “IOPscience,” Journal of Neural Engineering, https://iopscience.iop.org/article/10.1088/1741-2552/ab1e63 (accessed Oct. 20, 2023).
4. “Offline sorter,” Plexon,

https://plexon.com/products/offline-sorter/ (accessed Oct. 20, 2023).