Machine Learning Approach to Identifying Neural Features That Predict Rodent Behavior

Requirements and Specifications

Washington State University - Department of Psychology

Nunnerson Computing

Aidan Nunn, Charlie Nickerson

9/21/22

TABLE OF CONTENTS

I. Introduction

This will describe in detail the desired requirements and specifications of the Nunnerson Computing Project. Nunnerson Computing is tasked with designing a machine learning model that can predict imminent alcohol drinking in male and female rodents addicted to alcohol. Nunnerson Computing is also charged with creating a protocol that can assist future researchers in utilizing this model even if they have no experience in coding.

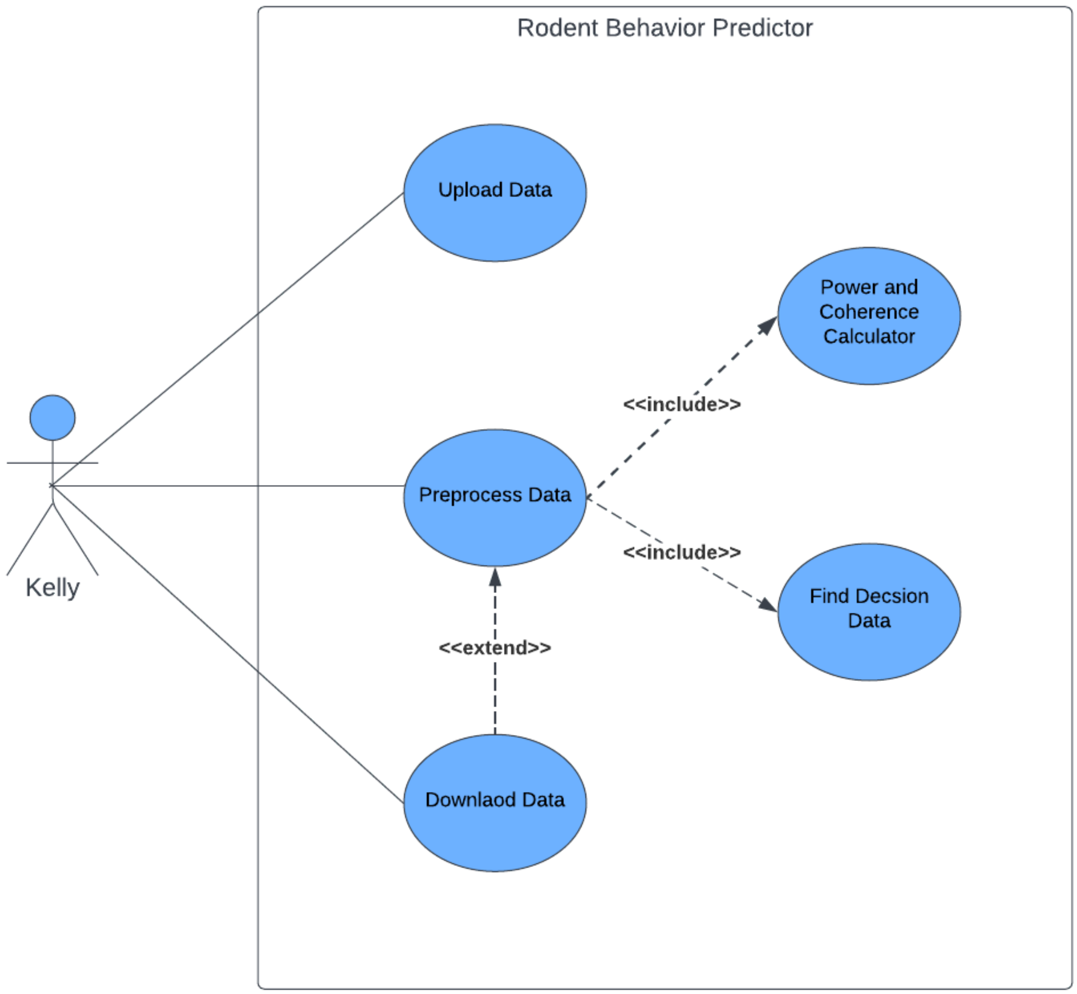
The purpose of this project is to help Angela Hendricks and her team of graduate and undergraduate researchers with speeding up their data collection process. In the past Hendricks and her team have had to manually record video of rodents and record data based on when the rodents looked like they were in the decision making process of consuming alcohol. This project will streamline this process by automating data collection so it no longer needs to be done by a human. Additionally, the data collected from the rodents won't be based on what the researcher observed and instead be based on the statistical probability of the model thus making the data more reliable and less prone to human error.

II. System Requirements and Specification

II.1. Use Cases

Story: Kelly Hewitt is a neuropsychologist at WSU who is researching rodent behavior. Kelly is currently studying four groups of rodents: non-alcohol dependent males, non-alcohol dependent females, alcohol dependent males, and alcohol dependent females. For each group of rodents, Kelly measures the local field potentials in three parts of the rodent's brain; The medial prefrontal cortex (mPFC), the nucleus accumbens shell (NAcSh), and the central nucleus of the amygdala (CeA). After collecting the data, Kelly would like to preprocess the data, so she only gets the LFP values measured when the rodent was in the decision-making process of drinking alcohol. To do this, Kelly opens our desktop application.

Once the application is opened, Kelly uploads LFP data for male alcoholic rodents into the app. Kelly then uses the program to clean the data by converting the LFP values to power and coherence values. Our machine learning model will then take these values and return the start and end timestamps of the recorded data, along with the power and coherence values derived from the LFP values recorded during this time for each rodent in the group. The timestamp outputs represent the period in which a rodent considers consuming alcohol. After preprocessing the data, Kelly saves the outputted data onto her computer and repeats the same process on the remaining three rodent groups. Kelly then closes the desktop application and begins her analysis of the newly processed data.



Story: Angela Henricks is a researcher at WSU who is researching rodent behavior. Angela uses a machine learning model to predict the actions of rodents who are dependent on alcohol. However, the model has been giving low accuracy on its predictions and Angela would like to retrain the model using a larger data pool. In order to accurately train the model, Angela needs to process an enormous amount of data, enough that it would take a very long time on a desktop computer. Angela needs to use the Kamiak Cluster to process the data efficiently.

Angela begins by opening our desktop application. She then loads each data file she would like the model to learn from into the application. Once the files are loaded, she can give the application her login credentials for Kamiak and command the application to ssh into the Kamiak server. The job will be sent to Kamiak, and Kamiak will build the model. The model file will be returned and can be saved onto Angela's local PC for use in the future.

Story: Angela Henricks is a researcher at WSU who is researching rodent behavior. Angela uses a machine learning model to predict the actions of rodents who are dependent on alcohol. She has recently computed a new model and hopes that it will have better accuracy than the previous one. Angela can use our desktop application to test the accuracy of the new model.

To start, Angela loads the new model into our desktop app. Then, she selects the option to test the accuracy. She is prompted to upload her own sample data or use one that is preinstalled into the app. The preinstalled sample will be small enough that it can be used on a desktop computer, but not too small as to undermine the accuracy of the test. Regardless of the chosen option, the application will run the model against the sample and return its accuracy. Angela can then decide whether the given accuracy is appropriate and choose whether to rebuild the model using more data or test the model again against more data.

II.2. Functional Requirements

II.2.1. [Machine Learning Model]

Rodent Decision Making Recognition Tool: This machine learning model tool must be able to recognize when a rodent is in the decision making process of consuming alcohol. This model will be tested on female and male rodents that are either addicted to alcohol or are sober. This will be important for knowing when to start and end data collection.
Source: Dr. Hendricks and her team have requested that this be in the project
Priority: Priority Level 0: Essential and required functionality

II.2.2 [Automated Data Collection]

Local Field Potential Data Collection Tool: One the machine learning model recognizes when a rodent is in the decision making process, this tool must begin collecting local field potential data (LFPs) in order to calculate the coherence and power values taking place within the rodent's brain. This is important data for Dr. Hendricks research and is essential to the project
Source: Dr. Hendricks and her team have requested that this be in the project
Priority: Priority Level 0: Essential and required for functionality

II.2.3. [Graphic User Interface]

User Interface for Automated Data Collection of Rodent LFPs: This tool is important for assisting users with using the predictive model to collect data. Users should be able easily import data into the program. This program should then neatly output the LFP data with timestamps of when that data was recorded. This entire process should be intuitive for the user and easily accessible for people who have little experience with coding.
Source: Dr. Hendricks and her team have expressed their desires for having this function in the project.
Priority: Priority Level 1: Desirable Function

II.2.4 [Power and Coherence Calculator]

Power and Coherence Calculator: This tool will take in the recorded LFP data and convert it into coherence and power values. These values will be taken in 5 second chunks of time. These values are important because they will be used for testing and training data in our machine learning model.
Source: Dr. Hendricks has informed us that the power and coherence values have been shown to work well for training a predictive model in the past.
Priority: Priority Level 0: Essential and required for functionality

II.2.5 [Interface with the Kamiak Cluster]

Interface with the Kamiak Cluster: This functionality will allow users to easily utilize the Kamiak Cluster for building machine learning models. The desktop application will prompt the user for their WSU credentials and use those to ssh into Kamiak and send the requested job to the cluster, with the output sent back to the desktop app.
Source: Dr. Henricks has informed us that she would like us to use the Kamiak Cluster to speed up calculations.
Priority: Priority Level 0: Essential and required for functionality

II.3. Non-Functional Requirements

Easy to Use: The final product should be easy to use by future researchers who have little to no knowledge in coding.

Readable Code: Once this project is complete the codebase should be easy to understand. All the code should have comments to describe the functionality of the code so future developers of the product can add to and maintain the code.

Fast: The speed it takes to process the data and calculate the outputs should be relatively fast. To speed up processing time we will be using WSUs Kamiak supercomputer.

Extensible: The code should be easily expandable by future developers who want to add features.

Documented: Directions on how to use functions and features in the project should be available for future users so if they do get confused on how to use the program, they can refer to a guide.

III. System Evolution

We expect system development to be a big factor in the design of this project. It's highly likely that the Henricks lab will request the aid of future capstone teams to help with their research, so we need to make sure our code is accessible to future developers. Additionally, we need to write protocols for use and ensure strict documentation so that members of the Henricks lab have many tools available to them in the case that they want to maintain or expand upon this software. We want to use Python instead of the original project language Matlab. Python is a common language that's useful for machine learning and is more widely used than Matlab. This will make the software easier for future developers to maintain.

IV. Glossary

Kamiak: A high performance computer on the Washington State University campus. Built out of a cluster of local computers connected via a high speed network.

Local Field Potential (LFP): The result of electrical activity in cells (mainly neurons).

Machine Learning: The use of computer programs that can learn based on input data and make predictions based on future data.

Machine Learning Model: A data file that has been trained to recognize specific data patterns.

Matlab: A high level programming language commonly used by mathematicians and engineers.

Python: An open-source, high level programming language with many libraries useful for machine learning.

Power: Power is the measure of each signal in frequency bands.

Coherence: Coherence is the correlation in activity in each site of the brain.

Medical Prefrontal Cortex: A region in the brain responsible for planning complex cognitive behavior, and decision making.

Nucleus Accumbans: A major component of the ventral striatum and has long been thought to be a key structure involved in mediating motivational processes.

Central Nucleus of the Amygdala: A location in the brain the is responsible for mediating many aspects of fear and anxiety.

7 Requirements and Specifications