Sources

* <https://archive.ics.uci.edu/ml/datasets/eeg+database>
* <https://academic.oup.com/alcalc/article/51/5/507/2237006> overview on genetic studies of AUD
* <https://auditscreen.org/> On the alcohol use disorder identification test (AUDIT). Contains links to scientific literature validating/qualifying the test
* <https://www.verywellmind.com/alcohol-use-disorders-identification-test-69492#:~:text=It%20is%20one%20of%20the,detecting%20hazardous%20or%20harmful%20drinking.&text=The%20test%20contains%2010%20multiple,alcohol%2Drelated%20problems%20or%20reactions>. Gives a number of 92% for the accuracy of AUDIT (not sure where they get it from though, not scientific literature)
* <https://onlinelibrary.wiley.com/doi/epdf/10.1097/01.ALC.0000164374.32229.A2> accuracy of AUDIT
* <https://www.datacamp.com/community/tutorials/convolutional-neural-networks-python> how to use CNN

Part 1 Big Picture

1. Define the science question or objective (Nick)

* Our objective is to use EEG time series data for subjects exposed to three different types of stimuli to identify if a subject is an alcoholic .
* The stimuli are pictures selected from the Snodgrass and Vanderwart picture set
  + The different types of stimuli are:
    - 1 pictures
    - 2 matching pictures
    - 2 non-matching pictures

2. How will your solution be used? (Suyash)

* Our solution will be able to take in the readings from all 64 electrodes for any given kind of stimulus, and predict if the subject is alcoholic or not
  + Can potentially be used to:
    - Help determine if some people are predisposed to alcoholism
    - Help diagnose someone with an alcohol use disorder (AUD)

3. What are the current solutions/workarounds (if any)? (Nick)

* AUDIT ranges between 70-90% effective for detecting hazardous or harmful drinking.
* The problem with this is that the test relies on the patient answering questions, which can always be unreliable
* It would be better to have a more objective way to do this using EEG readings

4. How should you frame this problem (supervised/unsupervised, online/offline, etc.)? (Nick)

* Supervised learning
* Binary categorization

5. How should performance be measured? (Suyash)

* Given that this is a classification task, one can simply use accuracy as a metric

6. Is the performance measure aligned with the scientific objective? (Max)

* Yes the performance measure is aligned with the objective because we just want to classify a person as alcoholic or not. Accuracy score works for binary classification.

7. What would be the minimum performance needed to reach the scientific objective? (Nick)

* AUDIT ranges between 70-90% effective for detecting hazardous or harmful drinking.
* Hopefully do better than that

8. What are comparable problems? Can you reuse experience or tools?

9. Is human expertise available? (Nick)

* I can always ask my prof if we have a neurophysics question

10. How would you solve the problem manually? (Max)

* I don’t think it is possible to solve this problem manually with this data set. The manual way to do this would be to give the patient the AUDIT questionaire.

11. List the assumptions you (or others) have made so far.

12. Verify assumptions if possible

**Intro and background (Zooey)**

The risk of developing Alcohol Use Disorder (AUD) is widely theorised to be driven by genetic predisposition in addition to environmental and psychological risk factors for addiction. While the actual neurophysiology of AUD is not as well understood, exploring its effects on brain activity can be a helpful tool in mapping out its potential characteristic psychophysiological processes as well as evaluating a patient’s risk of developing AUD.

In this report we

**Description of the dataset**

**Testing**

Baseline accuracy from a simple neural network on the S1 obj stimulus is 58%.

Baseline accuracy from a simple neural network on the S2 nomatch stimulus is 63%.