MLE Project – EEG Data

Melanie Kruger - 661995596

Project Description

I chose the problem based on what data feels relatively "unexplored" to me, and I stumbled upon EEG (electroencephalography) data. In short, EEG is a technique to capture the electrical signals in the brain, which can allow viewers to see what brain activity is happening where. This can allow researchers to monitor conditions like epilepsy [1], predict potentially life-threatening events such as sudden seizures [2], and even biometric identification [3].

This project is interesting to me from what questions this data could answer. A github repository has a list of different EEG data which is interesting to me [4], and I would like to pursue this further to learn more about EEG data and what machine learning methods that can be used in classifying such large datasets.

EEG Dataset

It'd be fun to analyze whether or not EEG correlates with a genetic predisposition to alcoholism [5]. I am open to other suggestions since there are many datasets and I chose this one just out of personal interest. However, one of the main problems is that EEG signals are usually noisy, don't conform to typical distributions which makes them hard to process, and individual differences can lead to huge differences in EEG signals.

Noise Problems

To get rid of noise, it seems like some sort of a bandpass filter could work, such as a third-order Butterworth bandpass filter [7] or use the EEGLAB Matlab toolbox (which I will not be paying for). The common alternative is to use deep learning, which has been used by several research groups. Since I will probably end up using deep learning, I will hopefully not need to process the data too much.

Another problem is in data artifacts, which occurs when someone blinks, moves their eyes, and contracts their head muscles. To get rid of the eye artifacts, researchers have been using concurrently recorded electro-oculogram data (ie. recording electrical signals at the eyes) to do regression analysis. Doing individual/principal component analysis to separate out signals within the noisy signal is another solution. To fix the muscle artifacts, canonical correlation analysis can be used [6]. Of course, there are several other methods that, for brevity reasons, are not included.

Analysis

Time-frequency analysis seems to be commonly used, where convoluted neural networks or singular value decomposition can be used. Other methods are naïve Bayesian, decision tree, K-nearest neighbor, support vector machine, and random forest [6]. My plan is to try these in sequence until something results in decent classification.

Evaluation

For deep learning models, I'll currently evaluate them primarily by using a confusion matrix, accuracy score, maybe the ROC curve as well. Basically, just the metrics covered in class, and any others if I find any special ones.

Citations

1. EEG: Tests, what is it & definition. Cleveland Clinic. (n.d.). https://my.clevelandclinic.org/health/diagnostics/9656-electroencephalogram-eeg

2. Ein Shoka, A. A., Dessouky, M. M., El-Sayed, A., & Hemdan, E. E.-D. (2023). EEG seizure detection: Concepts, techniques, challenges, and future trends. Multimedia Tools and Applications, 82(27), 42021–42051. https://doi.org/10.1007/s11042-023-15052-2

3. Marcel, S., & R. Millan, J. (2007). Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), 743–752. https://doi.org/10.1109/tpami.2007.1012

4. Meagmohit/EEG-datasets: A list of all public EEG-datasets. GitHub. (n.d.). https://github.com/meagmohit/EEG-Datasets

5. Eeg database. UCI Machine Learning Repository. (n.d.). https://archive.ics.uci.edu/dataset/121/eeg+database

6. Chaddad, A., Wu, Y., Kateb, R., & Bouridane, A. (2023). Electroencephalography Signal Processing: A comprehensive review and analysis of methods and Techniques. Sensors, 23(14), 6434. https://doi.org/10.3390/s23146434

7. 42. Ma C., Zhang M., Sun X., Wang H., Gao Z. Dynamic threshold distribution domain adaptation network: A cross-subject fatigue recognition method based on EEG signals. IEEE Trans. Cogn. Dev. Syst. 2023;early access doi: 10.1109/TCDS.2023.3257428. [CrossRef] [Google Scholar]