Machine Learning Methods for Analyzing Depression and Alcoholism

UNIVERSITY OF CAMBRIDGE

Gauransh Dhruv Tandon

University of Illinois at Urbana-Champaign

Mentored by Giovanna Maria Dimitri, University of Cambridge

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

ABSTRACT

Depression, as it stands, is third leading contributor to the global diseases – whose early detection can only help prevent various global diseases in the future. Researchers, also suggest that depression boosts alcoholic abuse and alcoholic abuse can also add to the tendency towards depression.

Therefore, in this research paper, alcoholic patients as well as controlled patients are analyzed in order to find a link between depression and alcoholism through their brain. Since, the analysis is done on the brain, therefore, prospective of analyzing both types of neuroimaging techniques: anatomical and functional; are explored. Ultimately, for the course of this paper, publicly available Electroencephalogram (EEG) signal processing is chosen – which is for both alcoholic and controlled patients.

Finally, explored data visualization methods for understanding EEG signal processing data, computed power spectral density for training Neural Networks and calculated the error of the model which is 0.16179.

LITERATURE REVIEW

Examined various past literatures that have used ML from different data sets including MRI scans and fMRI scans in reaching conclusions about numerous methods. After it was clear, how finding free data sets and processing large MRI and fMRI scans - was really hard, we examined studies relevant with the data set used in this research.

On this poster, I am explaining only those - which were relevant for my project, rest literatures are explained thoroughly in my research paper.

1. Professor Mallikarjun H M and Dr. H.N Suresh from universities in Bangalore, India analysed EEG signals, analysed EEG signals, first by converting EEG data set to .edf and then finding the power spectral analysis which they fed as inputs in the MATLAB ANFIS & nprtool boxes. Since power spectral analysis is the well-established method for analysis of EEG Signals, therefore this paper seemed the most logical for my project.
2. Research Scholars – Gopika Gopan K., Neelam Sinha and Dinesh Babu J, at the IIIT-Bangalore, India, analyzed EEG time series using non-linear features since the non-linear dependencies of different parameters were associated with EEG. They used K-means clustering which helped them in improving classification accuracy almost 110%. Since this is an ongoing research, K-means clustering will be used for analysis as well.
3. Krisztian Buza and Júlia Koller, researchers at BioIntelligence Lab, Institute of Genomic Medicine and Rare Disorders, documented how the EEG dataset had signals which appeared as nearest neighbors of a lot of other signals. Thereafter, they argued to incorporate this observation for the classification of EEG signals. This suggestion can be used for future studies regarding EEG signal processing.
4. Dr. Guohun Zhu and his team of 4 more postdoctoral research fellows at University of Southern Queensland, Australia, explored the horizontal visibility graph entropy (HVGE) approach to evaluate EEG signals from alcoholic subjects and controlled drinkers and compare with a sample entropy method. The accuracy of classification with 10-fold cross-validation is a whooping 87.5 % which again suggests the use of 10-fold validation system in analysis. These results demonstrate that the HVGE method is a promising approach for alcoholism identification by EEG signals.

PROJECT

DATA SET

The data set used for analysis of alcoholic and controlled subject EEGs, was publicly available EEG dataset from the UCI machine learning repository. The collection contained a total of 11028 EEG signals recorded from 122 subjects. Out which 45 are controlled patients and 77 were alcoholic patients. Every signal was documented for a second through 64 electrodes at 256 Hz. Hence, every signal is a 64-dimensional time series of length 256 in the data collection.

METHODS

The processing and analysis of the EEG Data is done using R - where packages, such as, eegkit, pwd, neuralnet, e107, caret, glmnet and ggplot2, are used. Data is converted into a readable .csv file after using eegkit package's in-built function. Then, mean, standard deviations and standard error is taken out, so that, they could be used for creating visualization. Thereafter, standard deviations and mean of the EEG data is taken for computing power spectral density using psd package. The psd pakage is using sine multitaper for computing power spectral. A visualization is then used to compare multitaper spectrum with cosine-tapered periodogram. For this research project, we chose multitaper over periodogram. After that, data collected from the power spectral estimations is used for training neural networks and the neural network is trained in the traditional way. There are 100 hidden layers, sigmoid function is 0.01 and the definition of eme.p and eme.par comes from what R package psd uses to define uniformly-tapered spectral estimates of the mean of EEG Data.

Data Extraction
- EEG Files converted to .eeg

Visualization
- Multi-channel EEG time courses
- Spatiotemporal Component
- Single-Channel EEG Time
- EEG CAP

Analysis
- Mean Value and Standard Deviation
- Power Spectral Analysis of Mean and Standard Deviation

Machine Learning Method
- Neural Networks

RESULTS AND DISCUSSION

Neural Networks produced a model with an error of 0.16179. This helps showcase a link between alcoholism and depression using neuroimaging. I have constantly tried exploring more and more machine learning methods like Support Vector Machines, Logistic Regression, K means Clustering, K Nearest Neighbour as well as Random Forrest which couldn't complete by the time everything else had to be finalized, hence, there is scope for further work – which is ongoing.

Based on the shortcomings of previous data sets as I studied in the literature review section, potential for improvements is clear in both, ongoing studies as well as future studies for developing better predictive models for use in depression diagnostics. Data sets containing information on the different types of depression, and information correlating diagnostic data with those particular types, become increasingly beneficial as their size increases.

A global effort must be made to unify existing data sets and standardize their dimensions, so that models may be built on the combined collections of data from depression diagnostics around the world. This, theoretically, would result in more accurate predictive models.

VISUALIZATIONS

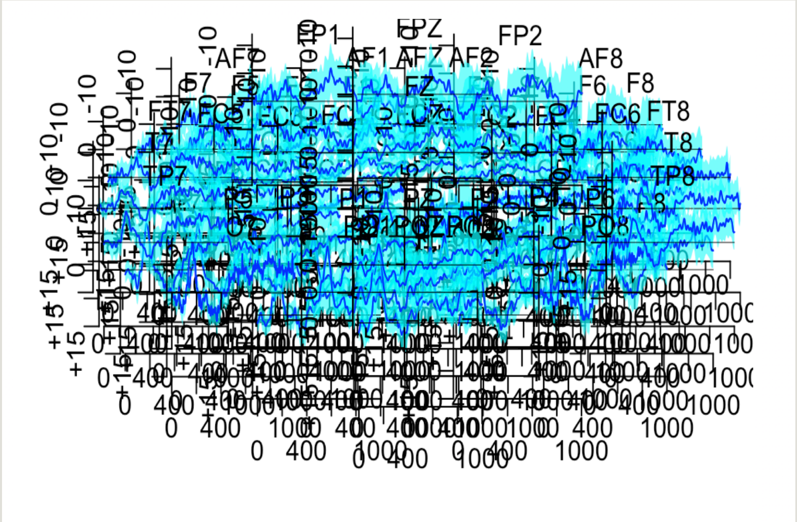

Figure 1: Multi-channel EEG time courses with subplots positioned according to electrode locations.

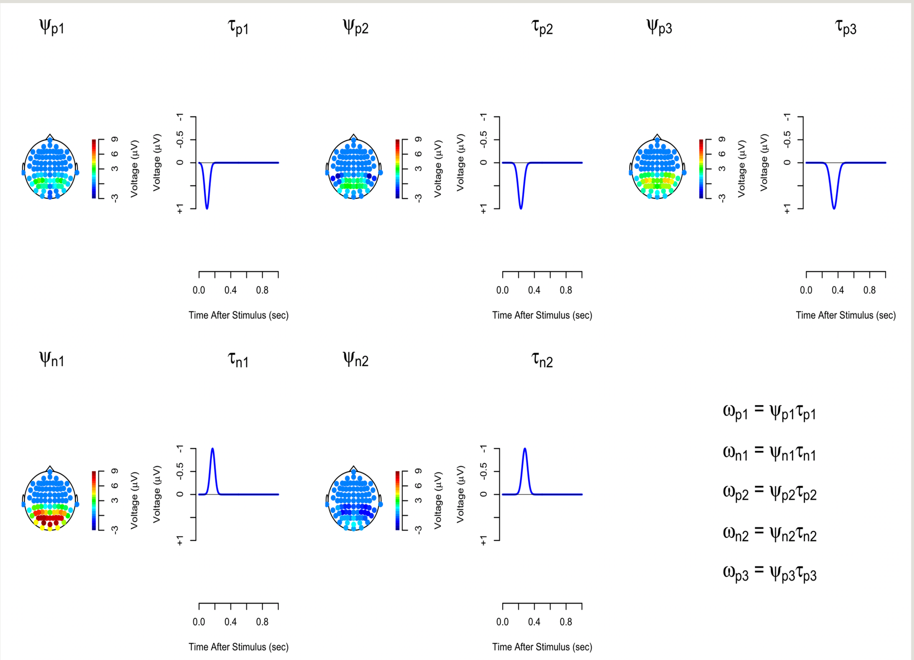

Figure 2: Spatiotemporal Component Functions. EEG Data is corresponding to the input channel(s), time point(s), coefficients and time shifts.

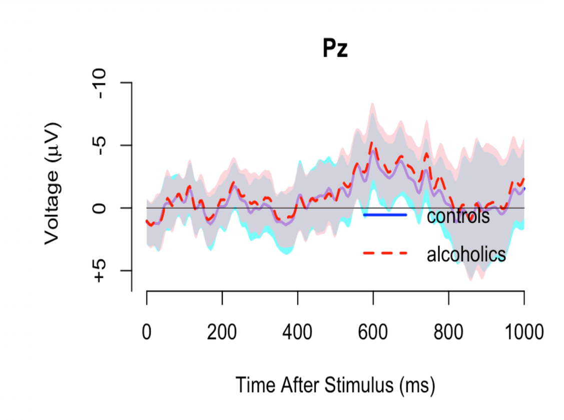

Figure 3: Single-Channel EEG Time course where standard error and mean are represented.

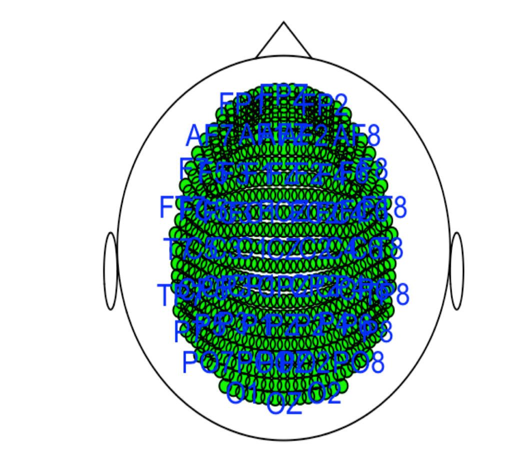

Figure 4: EEG Head Cap

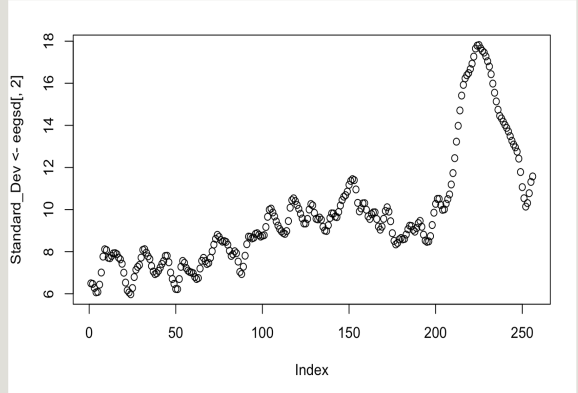

Figure 5a:Standard Deviation

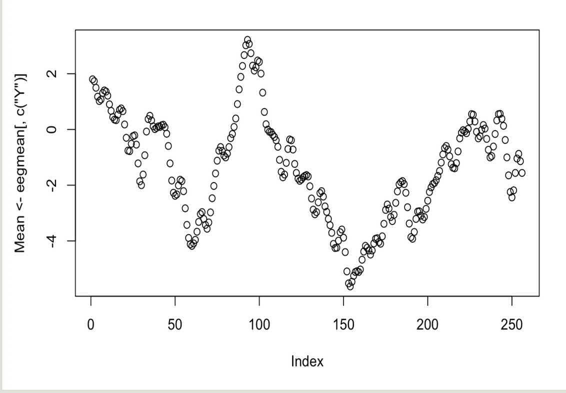

Figure 5b: Mean

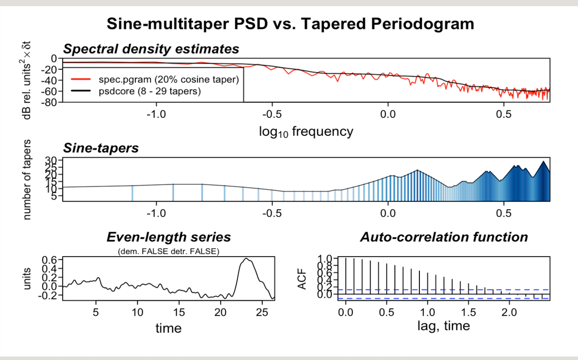

Figure 7:Comparison between Spectral Density Estimates from Multitapers and Peridograms for Standard Deviation

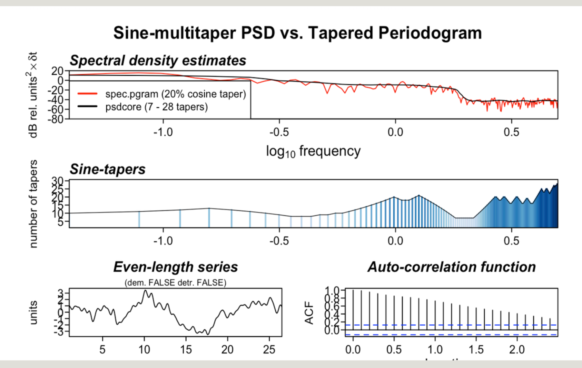

Figure 6: Comparison between Spectral Density Estimates from Multitapers and Peridograms for MEAN

REFERENCES

Mallikarjun, H. M., and H. N. Suresh. "Depression Level Prediction Using EEG Signal Processing." 2014 International Conference on Contemporary Computing and Informatics (IC3I) (2014) Web. 15 July 2018.

Buza, Krisztian and Júlia Koller. "Classification of electroencephalograph data: a hubness-aware approach." Acta Polytechnica Hungarica Vol. 13 - No. 2 (2016): 27–46. Web. 5 August 2018.

K. Gopika Gopan, et al., "EEG signal classification in non-linear framework with filtered training data." 23rd European Signal Processing Conference, France (2015): 624-628. Web. 8 August 2018.

Zhu, G., Li, Y. & Wen, P. "*Analysis of alcoholic EEG signals based on horizontal visibility graph entropy*." Brain Informatics 1(2014), 19–25. Web. 15 July 2018.

ACKNOWLEDGEMENT

I am really grateful of Ms. Giovanna Dimitiri for helping me through this paper as well as University Of Cambridge Computer Laboratory.