DISENTANGLING BRAIN ACTIVITY FROM EEG DATA USING LOGISTIC REGRESSION, XGBOOST, RNN AND CLASSIFICATION TREES.

Akshat Katiyar[1]

[1]*UG Scholar, Dept. of CSE, PSG College of Technology, Tamil Nadu, India*

Abstract - *EEG, is the method of choice to record the electrical activity generated by the brain via electrodes placed on the scalp surface. Electrodes are mounted in elastic caps similar to bathing caps, ensuring that the data can be collected from identical scalp positions across all respondents. This project aims at predicting a person's learning capability using EEG signals. The curiosity to learn about the changes that happen inside our brain while doing certain tasks drives this research. This research motivated us in developing a monitoring system which uses Electroencephalogram (EEG) as a fundamental physiological signal, to analyze and predict the learning capability of a person. The primary focus of this study is to identify the correlation between different data values extracted from raw EEG signal and how they change while doing different activities. This project disentangles brain activity through EEG signal data of 10 college students and predicts if a given student is confused or not while learning new unknown topics. The project aims at predicting the learning capabilities of an individual using EEG data from his brain.*

***Key Words*: Electroencephalogram, XGBoost, Logistic Regression, Classification Trees**

1. INTRODUCTION

An electroencephalogram (EEG) is a process used to monitor electrical activity of the brain. An EEG tracks and records brain wave patterns. Metal discs with thin wires are placed on the scalp, and then send signals to a computer to record the results. Normal electrical activity in the brain makes a recognizable pattern.

1.1 EEG DATA EXTRACTION

To extract the brain waves signals there are multiple neural waves recording devices in the market. For example, Mindwave mobile produced and marketed by NeuroSky. The devices like these are tiny in size and can be put on and removed easily and is connected to the software via Bluetooth. In this case MyndPlayer Pro Ver 2.3 by Neurosky was used as the software. The software can record the actual waveform of brain wave and the frequency of the various components of the waveform. The components are classified depending upon their frequency range. For an instance δ waves (0.5 Hz to 3 Hz), θ waves (4 Hz to7Hz), midrange γ waves (41 Hz to 50 Hz). In our case, we have 10 students who watch multiple 2-minute videos. We collect our experimenting data when the focus and the attention of the student is maximum. We remove the first and last 30 seconds. The different variations of the brain nerves during this activity is captured by the above-mentioned software.

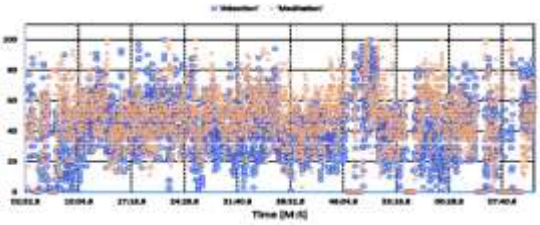


Figure 1.1 Variations of Brain Nerves

1.2 PREPROCESSING OF EEG SIGNALS

To interpret the signals that are fed into the computers the signals has to be processed. The EEG signals consists of several nonlinear distinct waveforms known as band components. The band components are complex and are categorized by the frequency range they fall into. The different band components are extracted from the raw EEG signals (Butterworth bandpass filters).

The following are the five primary bands of EEG signals:

- Delta (0.2–4Hz)
- Theta (4– 8 Hz)
- Alpha (8–13 Hz)
- Beta (13–30 Hz)
- Gamma (30– 55 Hz)

The small yet complex varying frequency structure found in scalp-recorded EEG waveforms contains detailed neuroelectric information. To analyze and isolate such waveforms or rhythms wavelet analysis is used. Since wavelet coefficients allow the precise noise filtering by attenuating the coefficients associated primarily with noise before reconstructing the signal with wavelet synthesis it has been used in the removal or separation of noise from the raw EEG waveforms.

There are two types of wavelet transforms: Continuous wavelet transform (CWT) and Discrete wavelet transforms (DWT). We in our experiment apply Discrete wavelet transform because it allows the analysis of signals by applying only discrete values of shift and scaling to form the discrete wavelets. The other advantage of applying DWT is that if suppose the original signal is sampled with a suitable set of shifting and scaling values, using the inverse of DWT the entire continuous signal can be reconstructed. The general Discrete wavelet representation is shown below:

$$\Psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}}\left(\frac{t - nb_0 a_0^m}{a_0^m}\right)$$

Where,

- integer's m and n control the wavelet shifting and scaling, respectively,
- $a0$ is a specified fixed dilation step parameter set at a value greater than 1,
- $b0$ is the location parameter which must be greater than zero.

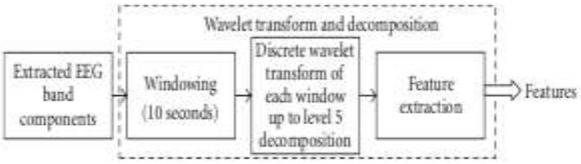

Wavelet transform and decomposition
Extracted EEG band components
Windowing (10 seconds)
Discrete wavelet transform of each window up to level 5 decomposition
Feature extraction
Features

Figure 1.2 EEG Decomposition and Extraction

The Figure 1.2 shows the EEG decomposition and extraction using the wavelet transformation. As shown in the figure each of the extracted band components are sent through the "Windowing step". In this step the interesting and boring portions of the band components based on the timestamps of the original EEG are extracted and sent through a windowing mechanism. In this mechanism, each band component signal is partitioned into tiny windows. After this step each window is decomposed using DWT. The decomposition process in wavelet transform can be performed iteratively into several levels. The values of the coefficients from the window interval is considered as the extracted features of EEG signals.

1.3 BRAIN WAVES CLASSIFICATION

Brain patterns form wave shapes that are commonly sinusoidal. Usually, they are measured from peak to peak and normally range from 0.5 to 100 µV in amplitude, which is about 100 times lower than ECG signals. By means of Fourier transform **power spectrum** from the raw EEG signal is derived. In power spectrum contribution of sine waves with different frequencies are visible. Although the spectrum is continuous, ranging from 0 Hz up to one half of sampling frequency, the brain state of the individual may make certain frequencies more dominant.

Brain waves have been categorized into four basic groups:

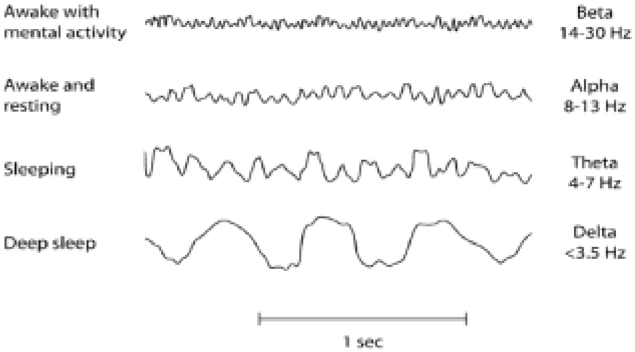
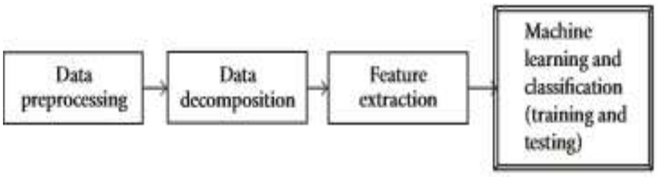


Figure 1.3 Different Categories of Brain Waves

2. SYSTEM DESIGN

The schematic diagram illustrates the overall method of this project. As shown EEG signals are acquired from the subjects during the experiment.



In the project we examine 10 students who watch educational videos of two kinds (Confusing one and the one which is from an easy topic). The duration of the video is 2 minutes and we take voltage difference of the waveforms emitted by the brains of the students.

Thus, the brain wave pattern is monitored using electrodes and sensors. After the activity of watching the video, students rate their level of confusion, either 1 or 0.'0' means the topic is not confusing and the value '1' means the topic was new or surprising to them.

The entire video of 2 minutes is cut into 1 min (by removing the first 30s and last 30s). And the response from students is taken for each 5s session. Thus, getting 12 modes for one video for one user.

We have two .csv files as our datasets in this project. The first one gives us the demographic information about the students and it is not much of use to us, since we perform our operations and apply our algorithms on the second file, which gives us the EEG signal information.

We initially perform our description analysis on this dataset by viewing the number of rows and columns present in it and also checking for missing values if any. There are actually 15 columns out of which 4 of them (student id, video id, user defined label, pre - defined label) are converted into integer for us to group and analyze the dataset easily.

The 4 described columns will also not affect the understanding of the particular subject for the student. So, they are included in the process of feature extraction. Since, all the features are of numerical data type (floating point numbers), we apply Pearson's coefficient as our tool. After applying, the Pearson's coefficient we find the relationship between the features by plotting them in different plots (Cross tables, bar graphs and heatmaps). Later we apply Logistic Regression, RNN, Boost and Decision Tree on the same.

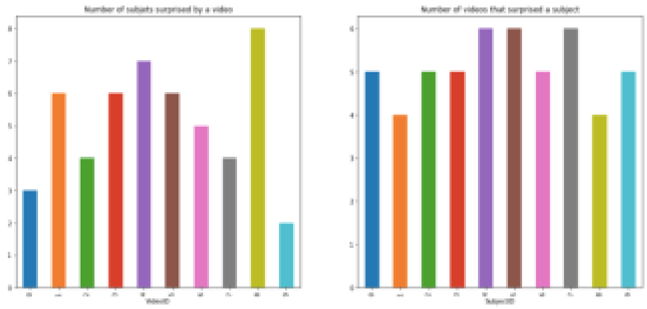
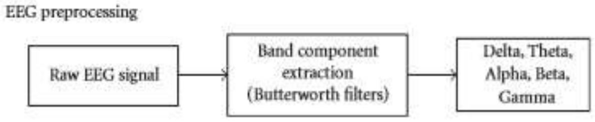


Figure 2 Comparison of Bar Graphs

We can infer from the above graph (Figure 3.a) that, video 9 was the easiest video since, only 2 students have told the video was surprising to them. In the same way, subject 1 has more learning capability than others since, he found only 4 videos were surprising.

2.1 DATA PREPROCESSING AND EXTRACTION


EEG preprocessing
Raw EEG signal
Band component extraction (Butterworth filters)
Delta, Theta, Alpha, Beta, Gamma

The EEG signal is comprised of a complex and nonlinear combination of several distinct waveforms which are also called band components. Each of the band components is categorized by the frequency range that they exist in. The state of consciousness of the individuals may make one frequency range more pronounced than others. As shown in Figure above, the different band components are extracted from the raw EEG signal using Butterworth bandpass filters. Five primary bands of the EEG signal are extracted, namely, Delta (0.2–4Hz), Theta (4– 8Hz), Alpha (8–13Hz), Beta (13–30Hz), and Gamma (30– 55Hz).

This Section tells about how key features are extracted from raw EEG signal. This Sample EEG data is of two subjects. It's given in a .EDF file which is basically Raw signal. The Output of such file is given below

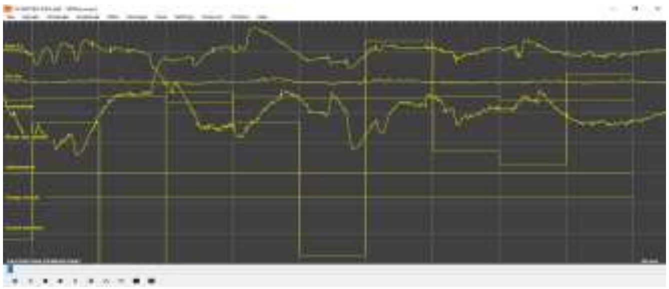


Figure 2.1 Raw EEG Signals

The above .EDF file representation (Figure 2.1) has 7 signals in which 2 are EEG channels.

2.2 FEATURE SELECTION

Feature selection is a process where you automatically select those features in your data that contribute most to the prediction variable or output in which you are interested.

Types of feature Selection performed

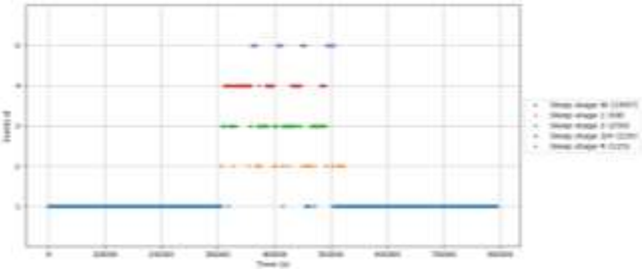
- Univariate Selection (operates only on positive variables)
- Recursive Feature Selection
- Pearson Correlation
- Feature Importance

2.2.1 UNIVARIATE SELECTION

Univariate Selection is a set of Statistical tests to select those features that have the strongest relationship with the output variable. Univariate Selection Works only on Positive Variable as we have a feature Raw which has negative values. The remaining 10 features are used for the classification. After Selecting the 10 features we use **SelectKBest and chi2** to find the best features among the selected 10. Later the most Dominant Features are used for Classification using Decision Tree, Logistic Regression, Extreme Gradient Boosting.

2.2.2 RECURSIVE FEATURE SELECTION

Recursive feature elimination is a feature selection method that fits a model and removes the weakest feature (or set of features) until the specified number of features is reached.



This method works by recursively selecting attributes and building a model on those attributes that remain. After selecting the 10 features we use **RFE** to find the best features among the selected 10.

2.2.3 PEARSON CORRELATION

When a data set has too many variables. Most of the variables are correlated. So, running a model on whole data returns poor accuracy. Finding the most positively correlated variables is mostly the best choice. Because they affect the output the most. Correlation is the mutual relationship between two features.

After Selecting the 11 features we use correlation (Pearson) to find the best features among the selected 11. Later the most Dominant Features are used for Classification using Decision Tree, Logistic Regression, Extreme Gradient Boosting.

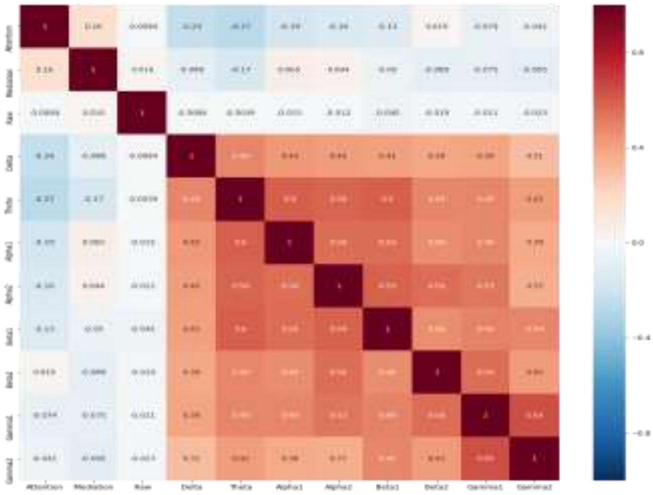


Figure 2.2.3 Pearson Correlation

Using Pearson Correlation, the heat map looks like this (Figure 2.2.3) it shows Delta, Alpha, Gamma, Beta, Theta are strongly correlated and are most dominant.

2.2.4 FEATURE IMPORTANCE

A common approach to eliminating features is to describe their relative importance to a model, then eliminate weak features or combinations of features and re-evaluate to see if the model fairs better during cross-validation this is feature importance. This method gives the feature importance of each feature of the dataset by using the feature importance property of the model. Feature importance gives the score for each feature of your data, the higher the score more important or relevant is the feature towards the output variable. After Selecting the 11 features we use **ExtraTreesClassifier** to find the best features among the selected 11. Later the most Dominant Features are used for Classification using Decision Tree, Logistic Regression, Extreme Gradient Boosting.

2.2.4 RAW SLEEP EEG DATASET

The Sleep Physio net dataset is annotated using 8 labels <physionet_labels> Wake (W), Stage 1, Stage 2, Stage 3, Stage 4 corresponding to the range from light sleep to deep sleep, REM sleep (R) where REM is the abbreviation for Rapid Eye Movement sleep, movement (M), and Stage (?) for any none scored segment.

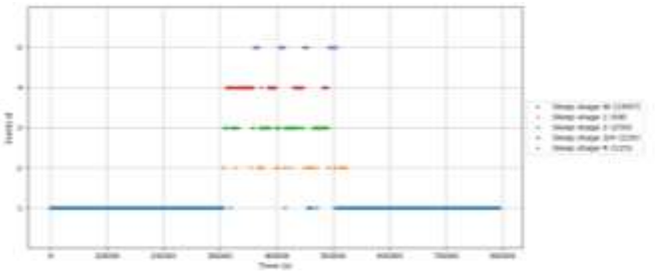
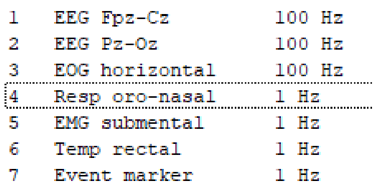


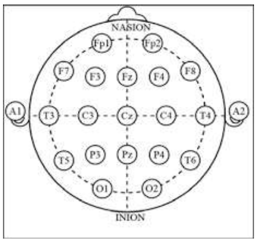
Figure 2.2.4 Sleep Stages

We will work only with 5 stages: Wake (W), Stage 1, Stage 2, Stage 3/4, and REM sleep (R). Here Stage 1, Stage 2, Stage 3/4 they are nothing just the sleep concentration where Stage 1 has less sleep concentration than Stage 2.

This is the Sleep data collected from a person (16 hrs. data). It has following values.

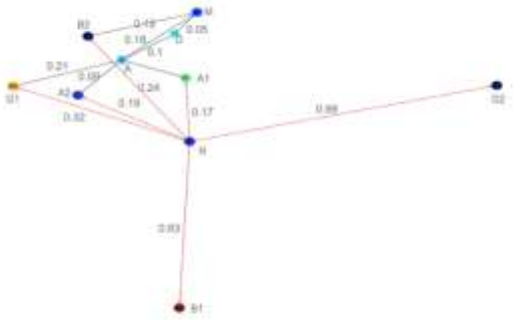


```
1   EEG Fpz-Cz        100 Hz
2   EEG Pz-Oz         100 Hz
3   EOG horizontal    100 Hz
4   Resp oro-nasal    1 Hz
5   EMG submental     1 Hz
6   Temp rectal       1 Hz
7   Event marker      1 Hz
```



Where FpZ (Intersection between Fz and Fp1 and Fp2 in the figure), Fz, Cz, Oz is just electrode labelling

2.2.5 DEPENDENT FEATURES(GRAPH)



Here each abbreviation means.

A-Attention, M-Meditation, R-Raw, D-Delta, A1-Alpha1, A2-Alpha2, B1-Beta1, B2-Beta2, G1-Gamma1, G2-Gamma2

The set of features dependent on each other is as follows:

MAIN FEATURE	DEPENDENT FEATURE
R	{A1, A2, B1, B2, G1, G2}
{A1, A2, G1, D, M}	A
{G2, A, D}	M

From the above graph following inferences are made. A (attention) is most dependent among all, R (Raw) is the origin of all-important feature like alpha, gamma etc. Hence, we conclude Attention is the most dominant feature in our study.

3. RESULT

After the code was completed various parameters were tuned for the best results. Adding all the features provided more accuracy than with using selected features. Classifiers were also changed to see which gives the best result. Here is the list of Classifier tried and the corresponding accuracies obtained with their corresponding run-time. Classification is one of the methods used to identify the certain pattern from feature extraction such as amplitude of EEG data, power spectral density (PSD), and Band power (BP). The performances of identify the pattern are depend on both features and classification algorithm. There are many classification algorithms used in the EEG analysis. Here following classifiers are used.

3.1 COMPARISON

- Better Dataset with high amount of training data.
- Independent features at least one so that all features are not needed for classification.
- Adding two new columns specifying the level of difficulty to make predictions more accurate.

The Figure 3.1(a) shows 3 algorithms along with their run-time, clearly taking all features gives the best accuracy but it does take a lot of time. Also, you can see XGBoost takes more time than the other two because it keeps searching for local minima and then finds a common global minimum. Hence it takes more time than other two classifiers.

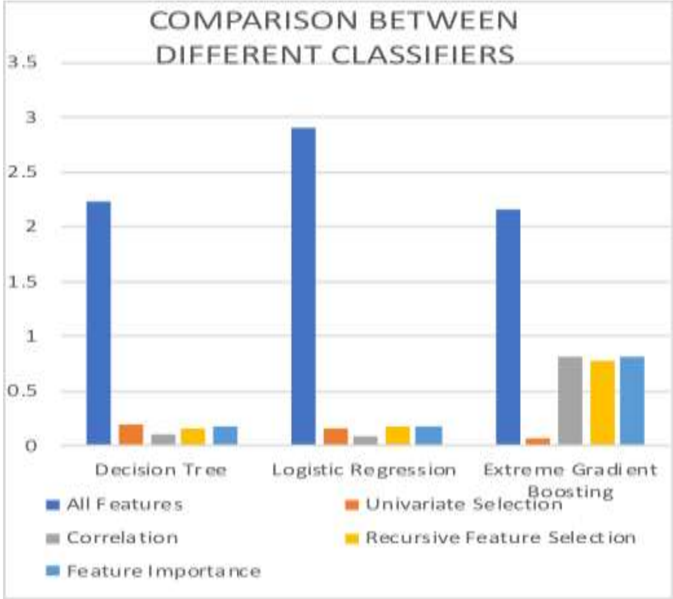


Figure 3.1(a) Comparison of Classifiers

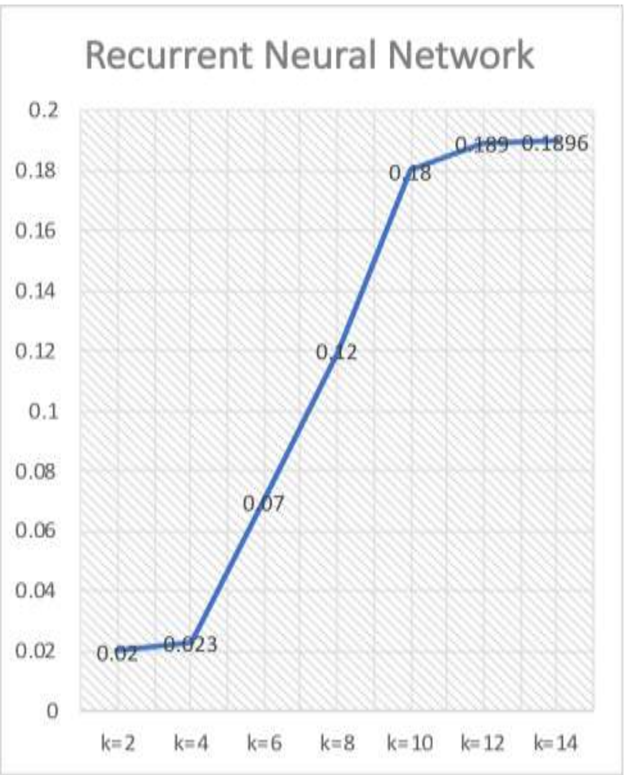


Figure 3.1(b) RNN Accuracy

The Figure 3.1(b) shows the accuracy of RNN. As seen after taking hidden layers greater than 10 the model goes to a saturation state and the accuracy remains same.

3. CONCLUSIONS

Predicting the learning capabilities of an individual is not an easy task as it has lot of complex tasks involved. It can be seen that reasonable amount of accuracy achieved when all the features are selected. Amongst the various machine learning models investigated the classification using **Extreme gradient boosting** seems to perform best (63.6%) on the combined feature set. However logistic regression using all features and recursive feature selection gives the closest accuracy to the best (61.3&, 61.2%). Also, by applying various feature selection algorithms on the dataset we come to know that we get the maximum accuracy of prediction when all the features are selected. This not only demonstrates that our dataset is small but also depicts that for making the prediction we can't depend only on our four important features (theta, alpha, beta, gamma). Although this work is limited to EEG data there is a room for improvement using clinical data and genetic data.

Here are some of the future work planned to improve the system's classification and prediction performance.

- A larger data set is needed to further validate this experiment. A larger data set is expected to provide a more robust classifier model.

- Having a more diverse base of features usually provides insight into some connate characteristics of the signal which might not be openly evident.
- Feature pruning and other classification methods need to be tried for increasing the accuracy.

Hence this project gives a little insight of how brain react to certain changes and how those changes can be used to predict certain tasks.

REFERENCES

[1] T. Blakely, K.J. Miller, R. P N Rao, Mark D.Holmes, and J.G. Ojemann, "Localization and classification of phonemes using high spatial resolution electrocortico graphy(ECoG) grids," in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, Aug 2008, pp.4964–4967.

[2] Spencer Kellis, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger, "Decoding spoken words using local field potentials recorded from thecortical surface," Journal of Neural Engineering, vol. 7, no. 5, pp.1–10,2010.

[3] Jess Bartels, Dinal Andreasen, Princewill Ehirim, HuiMao, Steven Seibert, E. Joe Wright, and Philip Kennedy,"Neurotrophic electrode: Method of assembly and implantation into human motor speech cortex," Journal of Neuro science Methods, vol. 174, no. 2, pp.168–176,2008.

[4] https://www.kdnuggets.com/2017/10/xgboost-top-machine-learning-method-kaggle-explaind.html

[5] https://physionet.org/physiobank/database/sleep-edfx/sleep-cassette