Names: Nitinram Velraj, Krishna Bairavi Soundararajan, Kaushik Iyer
42-675 Final Project Report:

Project Goal:

The focus of our project is to analyze electroencephalogram (EEG) signals to understand how the brain operates during various cognitive workload situations. This program can hopefully be used in a wide variety of applications in healthcare. Our project aims to classify cognitive workload, from a dataset of frontal EEG. Our project is based on a research paper that was focused on understanding whether short-term frontal EEG could be utilized as an adequate measuring tool for the mental workload. The paper is able to classify at a rate of 65-75% by utilizing an SVM model. Our aim is to get a better classification percentage, than the listed rate on the paper.

Background:

EEG is an electrophysiological monitoring method which is used to record the electrical activity of the brain. EEG works by measuring the voltage fluctuations resulting from ionic current within the neurons of the brain. There are various ways that an EEG experiment can be conducted, a standard, ambulatory, or video EEG [1]. The data that will be utilized in this project is going to be coming from a standard EEG test.

Upon the acquisition of data, the next part is understanding the captured data. These EEG signals can be broken down into one of the frequency based categories listed below [12]:

1. Infra-Low (less than 0.5 Hz)
2. Delta waves (0.5Hz to 3 Hz)
3. Theta waves (3Hz to 8 Hz)
4. Alpha waves (8Hz to 13 Hz)
5. Beta waves (13 Hz to 38Hz)
6. Gamma waves (38 Hz to 42 Hz)

Aside from the EEG, another aspect of the project is the machine learning(ML) methodology that is going to be utilized by Matlab to get a better understanding of the results. One of the things that we will be looking at is applying various models to indicate which model provides the best accuracy. While literature indicates that neural networks can provide the best accuracy, understanding the specifics of the various ML toolbox functions available is important. Understanding the manipulation that the functions can have on data is important in understanding what is the optimal function for not only the data set that is being reviewed but also future data sets.

Significance:

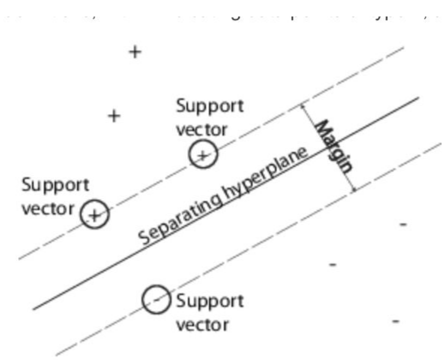
There is significant importance in using computational methods to better understand EEG datasets. Our EEG dataset comes from an experiment about mental workload, the data source is cited below. As discussed earlier, EEG waves can be broken down into different subcomponents. The processing in converting the raw EEG waves into these different types of waves is important because it can tell a lot about the patient. We use techniques from signal processing (specifical filters) to convert a raw EEG signal to a processed waveform. We later apply various statistical functions to the signals and arrive at our features. EEG cognitive workload study can be extended to study stress parameters for a patient. Cognitive workload study can also be extended to study the emotion parameters of subjects. We believe that this study can be extended to characterize the emotions and stress of special children who have impairments in communication and we wish to work in the project advancement in the future.

Aside from understanding emotion and stress, there are other applications that this work can be extended to. One of the first examples of where understanding the variations of data is important is analyzing drug efficacy. Since EEG allows for the scan of brain waves, ailments like seizures can be better monitored. With this monitoring, it can be analyzed which drugs are actually making an impact. Understanding this is important because it can help patients pick the medication that best suits them. On the other perspective, it also helps biomedical companies (Eg: Neurosky, Emotiv) create better products not only by understanding how their product compares to others on the market, but also identify new market gaps because of deficiencies in other products. Additional these computational methods will help in the real-time monitoring of a patient. This is significant especially in the biomedical space because you can pinpoint what exactly is happening, and the specifics of how a person is reacting.

Key Components For Computation:

- Signal Processing Analysis
    - Signal processing analysis was conducted on the dataset to filter out the noise/ undesirable artifacts based on specific frequency bands. The most common type of filter that was utilized throughout our code was the Butterworth filter. We are utilizing a bandpass Butterworth filter. The design of the filter is to have a frequency response that is flat. First a bandpass Butterworth filter with cutoff frequencies 0.01Hz and 42Hz as they were the minimum and maximum frequency components of EEG signals. We then proceeded to subdivide the EEG signal according to the ranges mentioned in the background by using bandpass Butterworth filters with corresponding cutoff frequencies. A notch filter is not used as the range of EEG signals is less the supply line frequency.
- SVM = Support Vector Machine

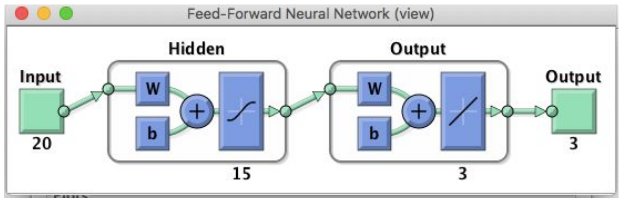
Figure 1: SVM Hyperplane



    - SVM is a type of supervised learning technique that splits the data into classes by finding a hyperplane. By identifying a distinguishing mark (hyperplane), the data can be separated. There are two parameters associated with SVM, the regularization parameter and gamma where gamma refers to the distance from the point and the hyperplane. SVM focuses on the classification of data by finding the best hyperplane that is able to separate the points of a respective class from another class. The respective hyperplanes that are being looked at depend on the number of dimensions since hyperplanes are the subspace n-1 planes. The diagram on the right demonstrates how the hyperplane is the divider between the + and - data points. Since our dataset has more than two classes, the multiclass SVM has to be utilized rather than binary SVM. In our code, we utilized the command “fitcecoc” to help train the multiclass SVM model.

- Neural Net
  - Artificial neural networks (ANNs) are computer models that were formulated by taking inspiration from the actual working of the human brain. These systems are able to “learn” by utilizing examples, generally without task-specific programming. An ANN consists of an input layer, a hidden layer(s) and an output layer. An activation (can be linear or non-linear) function is applied to the weighted sum of each previous layer to get the next layer (exception is the input layer). These weights are initialized before training and backpropagation algorithm is used to find the optimal weights based on the dataset. A neural network first traverses in the forward direction through the entire network and later travels in the backward direction and updates their weights.

Figure 2: Neural Net Model



- Features For Analysis
  - Mean
  - Standard Deviation
  - Root Sum of Squared Level
  - Peak to Root Mean Square (RMS) Level
  - Note: While other possible features were tried when writing the code, this group of four was the one that produced the best results. We tried other measures such as the interquartile range of the theta waves, median frequency, the skewness of theta waves in both frequency and time domain and kurtosis which is basically the peak of the frequency distribution. We also tried RMS, but RMS and Root sum of squared level are very similar and having both would be redundant so we removed that. From our manual trails of various features, we found the combination of these 4 features to give the highest accuracy so we ended up using them.
  - Manual Feature Selection
    - There were various combinations of features that were attempted manually and the combination yielding best out of the tested combinations was used to finally build the model.

Results:

Figure 3: Raw EEG signal

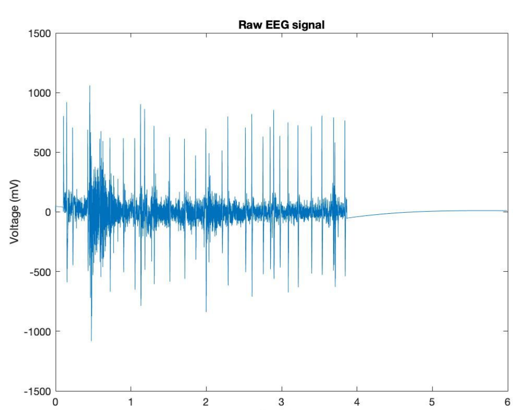


Figure 4: Power vs Frequency Plot showing components of EEG wave

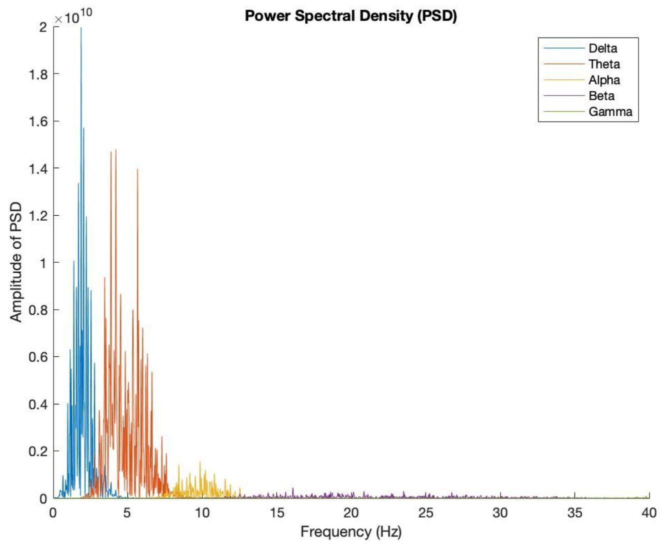


Table 1: Table Showing Accuracy Of Various Methods Over 1000 Runs

Method	Accuracy: Mean + Standard Deviation	Max Value
Neural Net	46.33 ± 5.39 %	60.13%
SVM	41.01 ± 4.37 %	46.15%

Conclusions:

The paper we used as the main reference [6] obtained an accuracy of around 65%. While peak scores got close to this value, the key number is the actual mean accuracy which is in the 40's. One of the possible reasons as to why we were not able to achieve this mark may be a consequence of using different features which do not capture the information in the EEG signals quite as well as the features used in the paper [6] has been able to. We tried to check how well other statistical measures capture information in EEG signals. Future works on trying out other features can be carried out to increase the performance of our model. The reason we tried neural networks is that with SVM you need to choose an appropriate kernel for optimal results. But with our features, the neural net did not perform very well. It is our belief that neural networks would yield better performance with the features mentioned in the paper as they seem to capture information in EEG signals better.

Most Intensive Element:

During our work on the project during the semester, the aspect of the code that ended taking the most of the time was determining the feature selection. We each were working on trying to come up with a combination of features that produced high accuracy results. This ended up being a lot of time researching various features that would be useful, and then implementing and then testing various combinations. While our accuracy isn't up to the paper [6], the combination that we put forth was the product of a lot of time and effort from initial accuracies of less than 35%.

References:

1. Ficker, David, and Maureen Gartner. "Electroencephalogram (EEG)." *EEG Electroencephalogram*, Apr. 2016, mayfieldclinic.com/pe-eeg.htm
2. Rager, Riki. "Measurement & Application & Overview." *ASET*, https://www.aset.org/Files/Public/10-20_System_Demonstration.Pd
3. Technologies, Trans Cranial. *Trans-Cranial 10/20 System Positioning Manual*, 2012, www.trans-cranial.com/local/manuals/10_20_pos_man_v1_0_pdf.pdf.
4. Sky, Neuro. "Greek Alphabet Soup – Making Sense of EEG Bands." *NeuroSky*, neurosky.com/2015/05/greek-alphabet-soup-making-sense-of-eeg-bands/.
5. "Biomedical Signals Acquisition (EEG)." *EEG*, www.medicine.mcgill.ca/physio/vlab/biomed_signals/eeg_n.htm.
6. So, Winnie K. Y., et al. "An Evaluation of Mental Workload with Frontal EEG." *PLOS ONE*, Public Library of Science, 2017, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0174949.
7. Patal, Savan "Chapter 2: SVM (Support Vector Machine) - Theory" https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72
8. https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#bsr5oqx
9. https://www.mathworks.com/help/stats/fitcecoc.html
10. "Artificial Neural Networks as Models of Neural Information Processing | Frontiers Research Topic". Retrieved 2018-02-20
11. Tatum, William O. (2014). "Ellen R. Grass Lecture: Extraordinary EEG". Neurodiagnostic Journal 54.1: 3–21.

Link To Original Dataset:

- https://figshare.com/articles/An_evaluation_of_mental_workload_with_frontal_EEG/4881464

Link To Google Drive With Data For part 1 (preprocessing):

- https://drive.google.com/drive/folders/1QlaDBY2ATue24rRAIepcHPSUx1coUVYk?usp=sharing

Link To Google Drive With Post Processed Data For Theta Waves only Has Been Renamed For The Classification Part of Program (classification):

- https://drive.google.com/drive/folders/1W4gdpdeHR4Zi1RuLLngGYfJQG2O5Z7p0?usp=sharing