A Real Time EEG Analysis System

Conference Paper · June 2016

DOI: 10.1109/DSC.2016.29

Yuezhe Li

UConn Health Center

Hong Lin

University of Houston - Downtown

A Real Time EEG Analysis System

Jonathan Garza, Yuchou Chang, Hong Lin
Department of Computer Science and Engineering Technology
University of Houston-Downtown
Houston, USA
linh@uhd.edu

Yuezhe Li
Department of Mathematics
Illinois State University
Normal, USA
yuezheli@gmail.com

***Abstract*— Electroencephalographic (EEG) data modeling is useful for developing applications in the areas of healthcare, as well as in the design of brain-computer interface (BCI). We built a system for brain state modeling, which includes a web server that can process uploaded electroencephalographic (EEG) data, store the data in a local database, and perform data analysis on the stored EEG data. This paper introduces a mobile application that is able to interact with the web server to render selected data and display analysis results from the web server. We aim to build an efficient self-adjusting brain wave modeling system that can seamlessly capture and analyze EEG brainwave data. The platform provides user friendly interface with secure data storage and analytics capabilities for wave analysis, statistical analysis, and categorical classification using a number of well-established machine learning algorithms. We also present a systematic method to understand how the variation of raw data sets used in training models affects the accuracy of machine learning algorithms, and then analyze the performance of machine learning algorithms under various computational implementations. Overall, the study describes a successfully built incorporated data analysis platform, and provides preliminary insights into the performance of common machine learning algorithms on the brain wave data sets.**

Keywords—Human-centered computing; Ubiquitous and mobile computing; electroencephalography (EEG); Machine Learning; Data Mining; Biomedical Informatics.

I. INTRODUCTION

Renowned scientist and philosopher Galvani was the first person to discover electrical activity in living organism in the 18th century [1]. Later, the electro physiologist Hans Berger successfully recorded electrical activity from the human brain using electroencephalography (EEG), which measures voltage oscillations due to ions flow in the neurons of the brain [1]. Today, EEG is one of the popular non-invasive techniques to record brain activity in clinical and research settings, and there is a wide range of applications for the analysis and interpretation of these measurements. EEG data carries an immense potential in its usability in various areas including human computer interaction, psychology, and neurological sciences. Therefore, it is a valuable endeavor to design an application that applies various analytical techniques to EEG data and predicts the state of the brain from which the data was acquired.

The EEG is also being used to develop innovative systems in healthcare and biomedical research. The recent study has been reported to discover links between emotional states of patients and their brain activity using machine learning algorithms [2]. The research analyzed EEG data collected during various emotional states from 40 Parkinson disease patients and healthy subjects using bispectrum feature and concluded that the higher frequency bands such as alpha, beta and gamma played important role in determining emotional states compared to lower frequency bands, delta and theta. In a different study, Direito and group have designed a model to identify the different states of the epileptic brain using topographic mapping relative to delta, theta, alpha, beta and gamma frequency [3]. The method achieved 89% accuracy in predicting abnormal vs normal brain states. These studies have reported the variability in analysis occur due to two major reasons, first based on feature extraction method implemented, and second the prediction of the model is directly proportional with the increase in the constant variables associated with the modeling equation. This overall underscores the complexity of applying mathematical models to a natural phenomenon such as brain activity [2,3].

A mobile application for comprehensive EEG data analysis is ideal for promoting applications using EEG tools in both physiological (e.g., clinical uses, sleep evaluation, fatigue detection, etc.) and psychological (cognitive sciences, BCI, etc.) scopes. The mobile application is part of an existing platform consisting of a webserver and database that the application interacts with and works together to perform analysis on EEG data for the user.

Nathan Holmberg, Burkhard Wunsche, and Ewan Tempero did a study on Interactive Web-Based Visualization in which they developed a framework to categorized different web-based technologies for 2D and 3D visualization [4]. DHTML which consists of HTML and JavaScript combined, performed well with disadvantages related to limited communication with servers at the time of the writing.

A study, done by Andrew V. Poliakov, Evan Albright, Kevin P. Hinshaw, et al., found that a major advantage, among others, in a server-client system setup is that the client's hardware does not need to be particular powerful as most of the processing of large data is done on the server side. Servers also tend to be more powerful than personal computers, even at the inexpensive end of servers. Another important factor with regards to server hardware is the popular inclusion of more than one CPU which makes it possible to run parallel data processing methods reducing the overall needed processing time of large amounts of data [5].

Another aspect of web-based system which has shown is that most users found a well-designed system to be easy to use and has a very quick learning curve [6]. This means that non-technical users can easily focus on the analysis of the data and less on the learning on how to use the system.

II. SYSTEM DESIGN

A mobile application has been developed to allow users to access the application from anywhere and in a convenient setting where a full computer may not be accessible. For the mobile application, the iOS platform was chosen and as such the objective-C programming language is the chosen language for the application. The mobile application contains an interface to the user to perform all the analysis methods hosted on the server.

A. The Web Server and the Interface

The web server is created with EEG analysis methods. These analysis methods have been created to work with the web interface to give web users a way to perform analysis on their uploaded research data. The analysis methods currently implemented on the web server are wave analysis, statistical analysis, and classification methods which utilize machine learning algorithms. In order to use these methods from a mobile app, an API is developed for communication with the web server in a convenient fashion. This API is created using the same language as the rest of the web server code, PHP. This language was chosen for its ease of use and popularity.

The mobile application contains a simple main menu view that lets the user select the option to perform analysis as seen in Figure 1. The user is currently presented with the option to perform analysis and this is the entry point for the application.

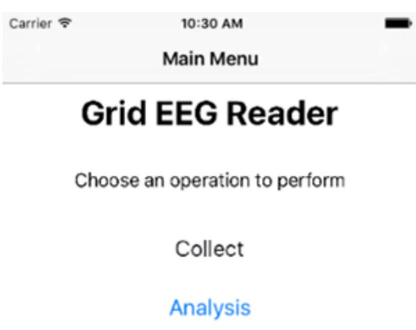


Figure 1. Application's Main Menu

This work is partially supported by the NSF UGI-CSTEM Grant.

The mobile application contains a view for selecting which analysis methods the user would like to perform. This could also be called an analysis main menu view as seen in Figure 2. The analysis main menu view allows the user to select an analysis method to perform. The analysis methods options are Wave, Stat., and Classification which are the same as on the website. After choosing their option, the user clicks select which loads an appropriate view to select options for that analysis method.

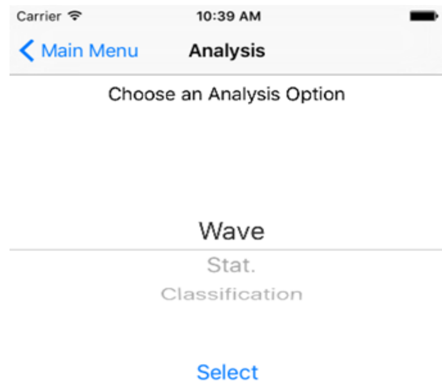


Figure 2. Analysis Main Menu.

B. Wave Analysis

For wave analysis, the user is presented with a different options view as seen in Figure 3. The user is allowed to select a table they wish to perform this analysis on and select which wave they wish to display. The waves can be selected in any combination and can be toggled on and off to make the selection. By default all the waves are selected. When the user has made their selections, they click the perform analysis button which will send the data to a wave analysis results view. That view will then send the request to the web server with the specified options and waits for a response. When it receives the response, it will parse the data received and display a wave graph with the selected data (Figure 4).

C. Statistical Analysis

For statistical analysis, the user is presented with the same options view seen in Figure 3. To allow the reuse of the options view, the chosen analysis method is stored as a property of the view. After the user selects their options and hits perform analysis button, the option’s view controller will send the selected options to a statistical analysis results view controller. The statistical analysis view controller will send the request to the web server similar to the wave analysis view controller with a difference being the analysis method value. Upon receiving a response, the view controller will parse the data into a format necessary for generating the bar chart to display on the graph. An example of this resulting bar chart can be seen in Figure 5. Each wave selected is displayed as its own set of bars on the graph. The blue bar shows the standard deviation of the EEG data and the red bar shows the mean of the EEG data. Although a quick glance can give the user an approximate view of the data, the user can click each bar to view the exact numerical value for that bar.

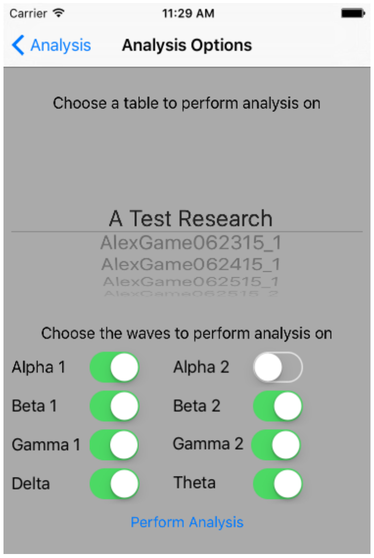


Figure 3. Wave Analysis Options View

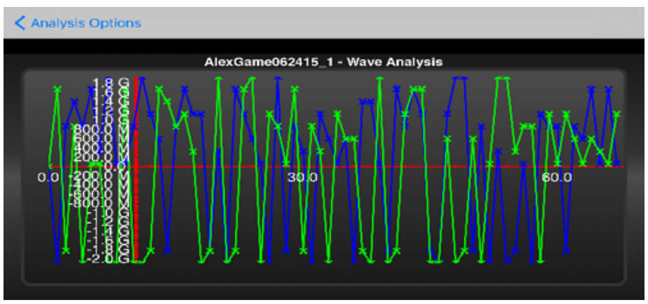


Figure 4. Wave Analysis View

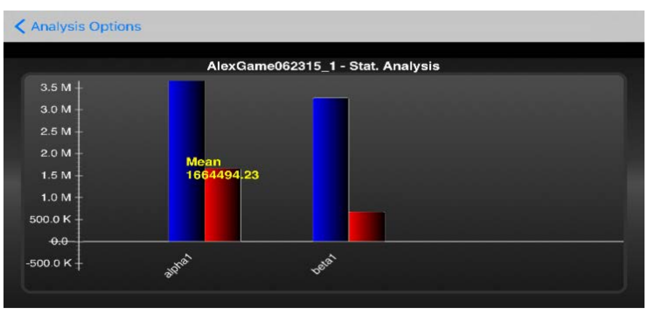


Figure 5. Stat. Analysis View

D. *Classification Analysis*

For the classification analysis method, after the user selects it as an option. They are present with a view to select options for this analysis method as seen in the top half of Figure 6.

The options view combines two views onto one scroll view, the top half never changing providing the user with the options for the classification algorithm to choose, and the classification mode to choose. When the user changes the mode, the second view changes to display the appropriate options for that mode. The previous view's entries are retained when the user reselects the other mode. The train mode is seen in Figure 7 while the test mode is seen in Figure 8. The train options provide the user the ability to select a list of tables for each class, class one and class two. The selection options are provided when the user clicks Select Data button for the class they wish to assign tables to. The list of tables is provided in their own table view. It also provides the user with the ability to set a name for the new model. For the test options, the user is presented with the options of selecting an old model which is appropriate for the selected classification algorithm. When the user changes the selected classification algorithm, the options for the old models change to display the new options appropriate for the new algorithm. The second option for the user to choose is a research data table to run the test on. If the user knows the list contains old data, they have the option to refresh the lists manually.

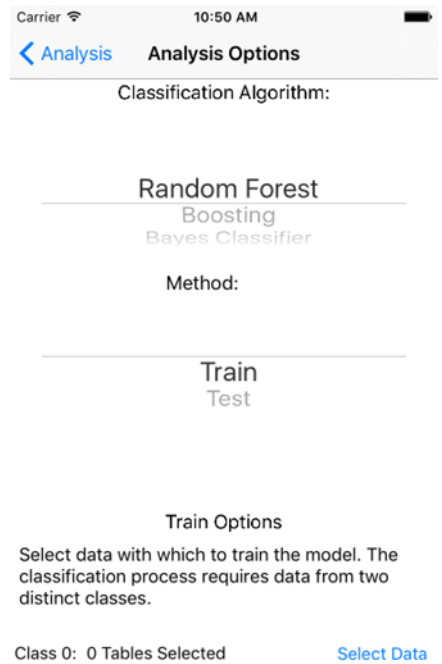


Figure 6. Classification Analysis Options View

After choosing their options, the user clicks train model or run test as appropriate to the mode. Both will send their options to a results view which will send their API request to the web server and displays the response from the server on a text area in the view. This can be seen in Figure 9. The response given back is dependent on the mode selected.

The results view displays an indicator while the mobile application waiting on a response from the server. Once the response is received and parsed, the indicator is hidden and the results displayed. From here the user can choose to go back to the previous view, the classifications options page with their selected options preserved or can choose to return to the home page and start a new analysis.

III. ON THE VALIDITY OF CLASSIFICATIONS

A statistical classification method is a useful tool in analyzing EEG data. Many modern machine learning algorithms and models have been successfully utilized in studying features hidden in EEG data collected from brains in different states. Here, by "brain state", we mean certain brain activity with some chosen features. For example, we may use "reading state" to mean the brain activities associated with reading. Supervised machine learning models include tree bagging, boost [7], random forest [8], and support vector machine [9]. Unsupervised machine learning algorithms, such as hierarchy clustering, are also utilized. Sample entropy, a method that aims to measure the uncertainty inside a sequence of data, also helps analyze brain activities through EEG records [10]. Different machine learning methods have different bases. Some are based on a decision tree; others are simply based on distance. Therefore, different classifiers exhibit different features. Our work shows that delta waves and theta waves change significantly between reading and meditation states.

Figure 7. Train Options

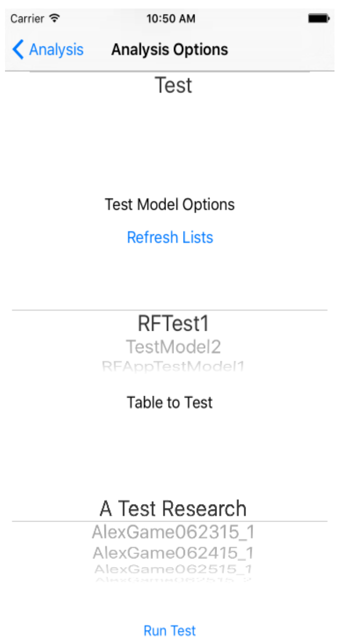
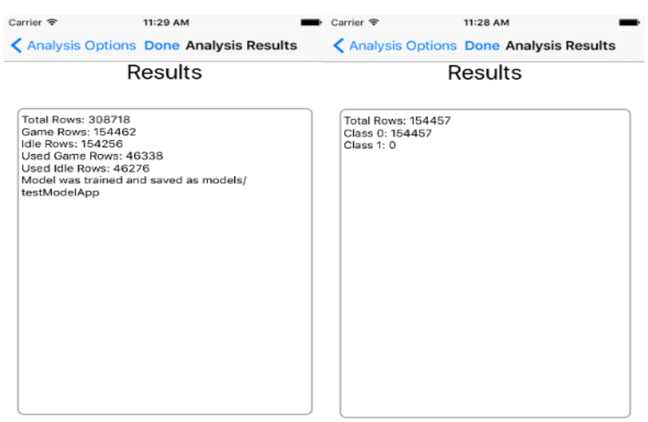


Figure 8. Test Options



(a) Train (b) Model

Figure 9. Results View

Our experiments have shown that using only delta waves and theta waves that k-nearest-neighbor classifiers have misclassification rates between 15% and 35%, and support vector machine classifiers have misclassification rates between 15% and 85%. These results indicate that using these two types of brain waves can have very good classification results, but sometimes they may not be informative enough [11-14]. Our experiments shows that besides theta and delta brain waves, that blink strength, another factor that can be measured by a Neurosky headset, also has a significant influence on brain state classification [15]. Based on all these result, we have designed an EEG data analysis system to classify brain states. However, it shows collinear it with different brain waves. As a consequence, when we build our brain state classifiers, we use brain waves and blink strength as mutually exclusive features.

We use data collected from Neurosky headsets. The data collected using Neurosky headsets is from 19 different subjects (volunteers). Neurosky headsets have a build-in system that can reduce the noise of the hardware and utilize embedded solutions for the signaling process and output [16, 17]. In our experiments, volunteers walk in, put on a Neurosky headset, and do what they are told to do (for example, playing a video game). Data is recorded in a txt file. We drop the first minute of data, as we believe brains need time to adjust. Then, for each subject, the recording time is between 5 and 10 minutes. Thus, for each subject, every data entry is between 153600 and 307200 (sample rate 512 times per second). Some subjects have the EEG data collected multiple times. The EEG data of each subject (volunteer) was acquired at different times and in 5 different states. Our classification models are built based on tree bagging, random forest, k-nearest neighbors, boost, and support vector machine.

Our result shows that with the increasing size of training data, misclassification rate decrease, even though for some algorithms, such as tree bagging or boost, the decrease is not significant (Figure 10). Their tree like classification and multiple time of iteration make them robust that guarantee that even if the data set is small and obscure, they are able to give low misclassification rate.

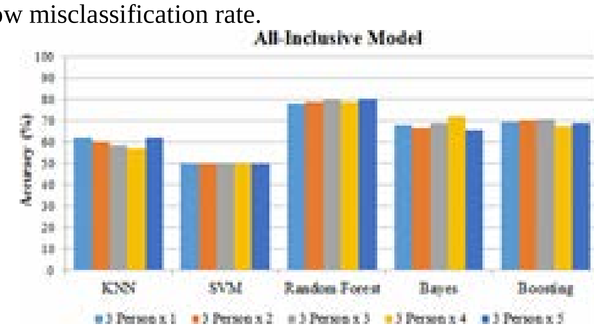


Figure 10. Misclassification of different algorithms with different input training data size

On the other hand, with the increasing training data, regardless due to increasing training data from a specific subject, or due to the increasing amount of training subjects, misclassification of support vector machine, which resembles classification algorithms based on distance, decreases (Figures 11, 12). Note that Figure 11 shows a significant decrease misclassification rate in support vector machine, while bagging and boost even increase a little. The result of the majority votes stays the same. We can expect a lower misclassification rate in majority vote with the greater amount of training data. Majority votes method combines both advantages and disadvantages of these three classification algorithms. There is no guarantee that majority votes always give the best classification result, but in generally, it is robust when the data size is small, and shows decrease in misclassification rate when training data size grows bigger (Figure 11).

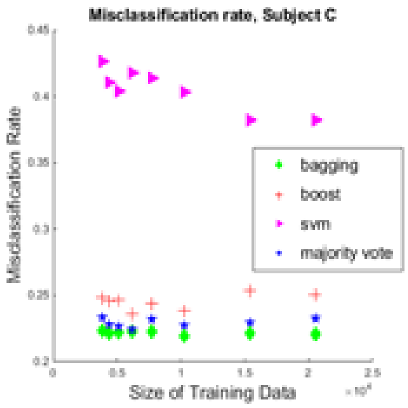


Figure 11. Misclassification of 4 different classification algorithms using data collected from subject C.

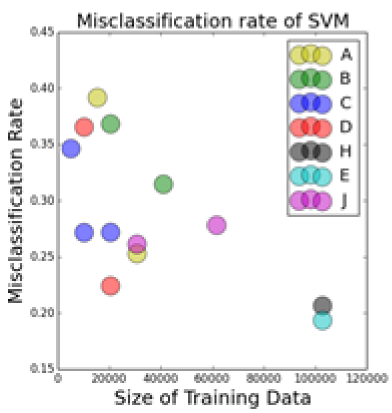


Figure 12. Misclassification rate of support vector machine on data collected from different subjects. It shows a significant decreasing trend while size of training data increases.

Our previous research [12] also shows that even with less waves, we can still have a good classification result using support vector machine (Figure 13). There we only use two waves, delta wave and theta wave. The result shows that there is not a lot mixing between different brain states, consequently we have a low misclassification rate.

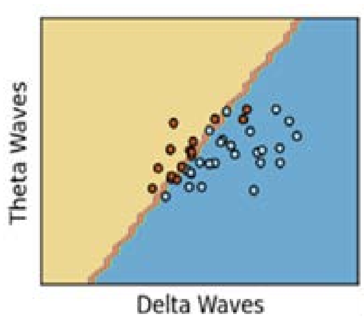


Figure 13. Support vector machine boundaries.

We also test how number of channels would influence performance of Gaussian mixture model based on sample entropy. In this scenario, we calculate sample entropy of our training and test data. We normalize our data before we calculate sample entropy of every minute's amplitude entropy. After calculation of entropies, we perform our classification method, Gaussian mixture model, on the entropies. Entropies are randomly split into two sets, to train and to test. We use training dataset to train our classifier, and then classify test dataset. We compare the classification states with the true states in our test dataset and then calculate the misclassification rate.

We use two groups of data here. Group 1 is EEG datasets collected from one channel, group 2 is EEG datasets from 7 channels. Group 1 gives misclassification rate of 0.33, while group 2 gives misclassification rate of 0. We exhibit entropies of group 2 in Figure 14, where 7 labels on the diagonal line are the names of channels. Filled blue dots are for talking state, hollow red diamonds are for idle state, and filled green triangles are for meditation state. It is clear that these 3 brain states do not overlap significantly. This indicates that their entropies are significantly different. Consequently, entropy is a good method to extract features from data.

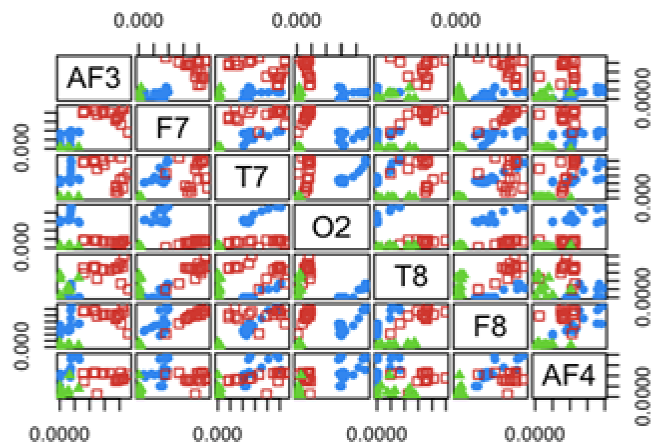


Figure 14. Sample entropies of group 2 datasets.

The significant decrease in the misclassification rate also demonstrates that increasing number of channels improves classification result. This is to be expected. It is widely agreed that different brain areas behave differently in a certain brain state [11, 12].

IV. CONCLUDING REMARKS

The design of the mobile application allows that user to perform all the analysis visualizations efficiently and natively, in addition to the options to perform them on the research website. Due to the limitation on the computational power of mobile phones and the constrains in maintaining decent response time to user requests, the analysis computations are performed on the web server with only the visualization of the results needing to be generated on the mobile application. This is important as some analysis methods require lots of processing which could take long on the limited processing power of the mobile device.

A data collection feature is needed for the mobile application. During the course of this project, the data collection method was under development due to the fact that different EEG headsets are used in data collection, and different EEG headsets collect data with different characteristics and different formats. By adding a data collection feature to the mobile application, the user could perform both operations of research, data collection and data analysis.

The mobile application, as well as the web server and the data storage and analysis system, presented in this paper will allow users to view and analyze EEG data conveniently through a mobile phone. This application paves the way to a system that helps users self-monitor their own brain states in a real time fashion, and such a system can accommodate vast possibilities of surveillance services for various kinds of patients in healthcare systems.

ACKNOWLEDGMENT

This work is partially supported by the NSF UGI-CSTEM Grant. The authors of this paper give thanks to our other team members who have helped in the development of the mobile apps. Special thanks to give to Gregor Schreiber and Minghao Yang, who developed the earlier versions of mobile apps [15] for the brain waves analysis project. Christopher Early and Alexander Chan contributed to the implementation of the brain waves analysis methods on the web server [13].

REFERENCES

[1] J Kropotov, Quantitative EEG, Event-Related Potentials and Neurotherapy. 525 B Street, Suite 1900, San Diego, CA 92101-4495, USA: Elsevier Inc, 2009.

[2] R Yuvaraj, M Murugappan, N Ibrahim, K Sundaraj, M Omar, K Mohamad, and R Palaniappan, "Optimal set of EEG features for emotional state classification and trajectory visualization in Parkinson's disease," International Journal of Psychophysiology, Vol. 94, Issue 3, December 2014, pp. 482-495.

[3] B Direito, C Teixeira, B Ribeiro, M Branco, F Sales, and A Dourado, "Modeling epileptic brain states using EEG spectral analysis and topographic mapping," Journal of Neuroscience Methods, Vol 210, Issue 2, September 2012, pp. 220-229.

[4] N. Holmberg, B. Wunsche, and E. Tempero. "A framework for interactive web-based visualization"; *AUIC '06 Proceedings of the 7th Australasian User interface conference* (Australian Computer Society, Inc.), 50, 137-144, Jan. 2006.

[5] Andrew V. Poliakov , Evan Albright , Kevin P. Hinshaw , David P. Corina , George Ojemann , Richard F. Martin , James F. Brinkley. "Server-based Approach to Web Visualization of Integrated Three-dimensional Brain Imaging Data"; *Journal of the American Medical Informatics Association*, 12, 2, 140-151, Mar 2005, DOI: 10.1197/jamia.M1671.

[6] Lourenço, A., Plácido da Silva, H., Carreiras, C., Priscila Alves, A., & L. N. Fred, A. "A web-based platform for biosignal visualization and annotation"; *Multimedia Tools and Applications* (Springer US), 70, 1, 433-460, May 2014.

[7] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for eeg signal classification," Pattern Recognition Letters, vol. 28, no. 15, pp. 2157–2163, 2007.

[8] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single eeg channel and random forest classifier," Computer methods and programs in biomedicine, vol. 108, no. 1, pp. 10–19, 2012.

[9] I. Gu¨ler and E. D. U¨ . Beyli, "Multiclass support vector machines for eeg-signals classification," Information Technology in Biomedicine, IEEE Transactions on, vol. 11, no. 2, pp. 117–126, 2007.

[10] Y. Song, P. Li`o et al., "A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine," Journal of Biomedical Science and Engineering, vol. 3, no. 06, p. 556, 2010.

[11] H. Lin, "Core mind-body practices and scientific calibration," Journal of Man, Machine and Technology (JMMT), Vol. 1, No. 1, 2012, pp. 73-89.

[12] S. Subedi, Y. Li, C. Early, A. Chan, J. Garza, G. Schreiber, Y. Chang, and H. Lin, A System for the Analysis of EEG Data and Brain State Modeling. Elsevier/MK, 2016.

[13] C. Early, A. Chan, S. Subedi, Y. Li, and H. Lin, "Analysis of machine learning algorithms on eeg data for brain state intelligence," in 2015 Workshop on Health Informatics and Data Science (HI-DS). IEEE International Conference on Bioinformatics and Biomedicine, 2015, pp. 793-799.

[14] S. Subedi, J. Garza, Y. Cui, and H. Lin, "Modeling of eeg data for brain state intelligence and ehealth," 9th China Summer Workshop on Information Management (CSWIM 2015), Hefei, China, June 27-28, 2015,

[15] G. Schreiber, H. Lin, J. Garza, Y. Zhang, and MinghaoYang, EEG visualization and analysis techniques. Springer book series Health Information Science, 2015.

[16] K. Crowley, A. Sliney, I. Pitt, and D. Murphy, "Evaluating a brain-computer interface to categorise human emotional response," in 2010 10th IEEE International Conference on Advanced Learning Technologies. IEEE, 2010, pp. 276–278.

[17] C.-J. Lin, C.-H. Ding, C.-C. Liu, and Y.-L. Liu, "Development of a real-time drowsiness warning system based on an embedded system," in Advanced Robotics and Intelligent Systems (ARIS), 2015 International Conference on. IEEE, 2015, pp. 1–4.