Phase-Lag Index vs. Coherence as Features for EEG Classification

Minsi Hu[1]

[1] University of Maryland, College Park, MD 20742, USA
mhu0315@umd.edu

Abstract. Background: Recently, significant attention has been drawn to the discovery and evaluation of novel methods of functional connectivity (FC). Specifically, we evaluated the performance of two FC methods, namely phase lag index (PLI) and coherence. We started by reading in the STEW raw EEG dataset and processing the data. Then, we computed PLI and coherence features and used those with a variety of ML algorithms to classify high vs. low mental work-load. Our results showed a significant difference between the efficacy of PLI vs. coherence features, where using PLI features alone produced an accuracy of ~65% while coherence features alone produced an accuracy of ~85% with multiple different ML models. However, there were multiple inconsistencies and limitations of the methodology, so further investigation is needed to determine the accuracy of these results.

1 Background

Brain-computer interfaces (BCIs) have long been seen as control interfaces that translate changes in brain activity [4]. BCI can be passive or active. A passive BCI interprets brain signals to give insight into the cognitive state of an individual [5]. One such example of passive BCI is interpreting brain signals in the form of EEG data to determine if a subject is under high or low mental workload. Mental workload reflects the amount of mental resources required to perform a set of concurrent tasks [6]. Sustained high mental workload will cause mental fatigue, decreased performance, and even detrimental health effects in the long run [6], giving us the motivation to implement passive BCI, particularly in a real-time context. However, in this project, we implement a software to simulate BCI in an offline context to evaluate the performance of different functional connectivity (FC) methods. Functional connectivity is a quantitative measure of how regions of the brain interact with each other, which we will use to classify between two levels of mental workload [7]. The functional connectivity methods we will be comparing are Phase Lag Index (PLI) and Coherence.

2 Methodology

The process at hand is the following: read in the STEW raw EEG dataset and clean/process the data. Then, compute PLI and coherence features and use those features to classify high vs. low mental workload. Finally, use those results to determine whether the PLI or coherence features were more effective for classification.

2.1 STEW Dataset

The Simultaneous Task EEG Workload Dataset (STEW Dataset) was used, which consists of raw EEG data from 48 male subjects [1]. EEG was recorded under rest and workload conditions, which we labeled with low and high mental workload respectively. In the rest condition, EEG was recorded when the subjects were at rest and not performing any tests [1]. In the workload condition, EEG was recorded while subjects were performing the Simultaneous Capacity test (SIMKAP)—a test designed to assess an individual's multitasking and stress tolerance [1]. The EEG was collected using Emotiv EPOC EEG headset with a sampling frequency of 128 Hz and fourteen electrodes located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 [1]. Both EEG recordings spanned 3 minutes, with the initial and final 15 seconds being excluded to minimize impact of between-task activity [1]. The result is two EEG recordings (one of low workload and high workload), comprising a total of 19200 samples for each of the fourteen channels, for all 48 subjects. Before feature extraction, the data was cleaned with a basic 1-50 Hz bandpass, 6th order Butterworth filter. Fig. 1 shows the raw EEG data for low and high mental workload from the STEW Dataset, while Fig. 2 shows the cleaned EEG data.

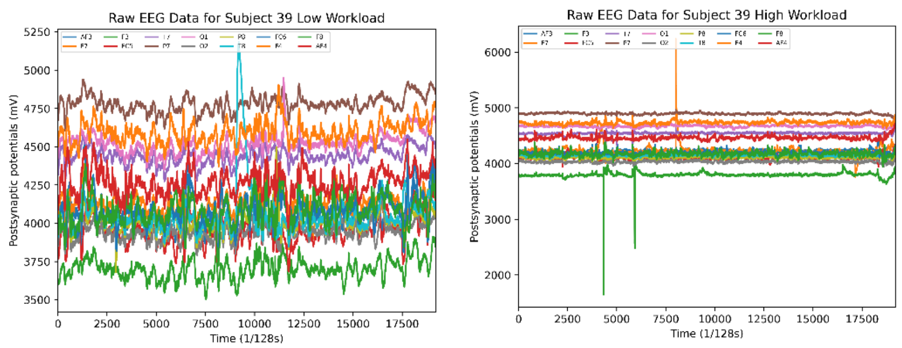

Fig. 1. Raw EEG for Low vs. High Mental Workload

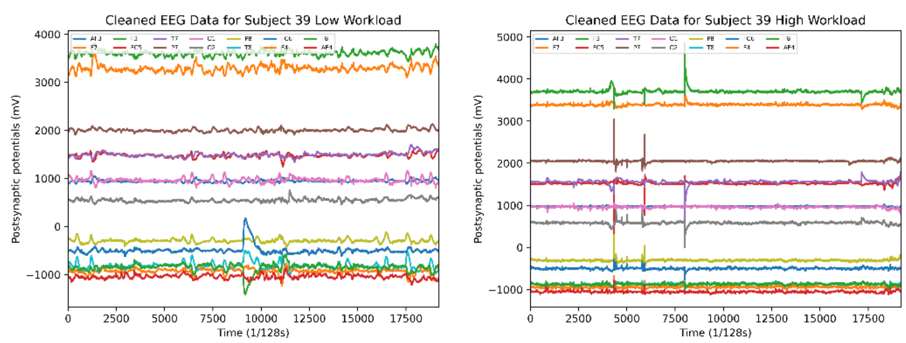

Fig. 2. Cleaned EEG for Low vs. High Mental Workload

2.2 Phase Lag Index

PLI (Phase Lag Index) is one of the features we used to classify between two levels of mental workload. The concept of PLI starts with the idea of an instantaneous phase, which is a measurement of the position/angle of a wave relative to its cycle at a specific point in time. Phase difference quantifies the difference between the phase values of two signals in a specific point in time, whereas phase lag characterizes the phase of a signal relative to the phase of another signal sharing the same period. PLI allows us to quantify phase lag by describing the consistency of phase differences between two signals across an entire time period.

PLI is defined as the following, where *n = number of trials*, *j = phase of signal X*, *k = phase of signal Y*, and *t = trial*:

$$PLI = \left| n^{-1} \sum_{t=1}^{n} sgn\left(Im\left[e^{i(\theta^j - \theta^k)_t}\right]\right) \right| \quad (1)$$

The $(\theta^j - \theta^k)_t$ term is simply the phase difference between signal X and signal Y at instantaneous time/trial *t*. The $e^{i(\theta^j - \theta^k)_t}$ term can be thought of as plotting the phase difference on the complex plane. $Im\left[e^{i(\theta^j - \theta^k)_t}\right]$ grabs the imaginary component of the plotted phase difference points which is positive when signal Y lags behind signal X, and negative when signal X lags behind signal Y—which can be thought of as instantaneous phase lags. Then, the signum (*sgn*) function is applied to every one of these points, returning -1 when signal X is lagging behind signal Y and 1 when signal Y is lagging behind signal X. Finally, we average all these values across the entire time frame and take the absolute value to get the PLI connectivity value between signal X and signal Y.

The resulting PLI will be a value between 0 and 1, where 0 indicates no consistent phase lag and 1 indicates perfectly consistent phase lag. Further, a PLI value of 0 indicates either perfect synchronization between two signals, or volume conduction—the same signal being picked up by two different sensors—which we generally do not want. A PLI value of 1 indicates all our instantaneous phase lags had the same sign, allowing us to say there exists a statistical dependence between two signals and thus they show connectivity.

For feature extraction, we calculate the PLI between all 14 sensors, giving us a PLI connectivity matrix. We use this PLI connectivity matrix as a graph adjacency matrix to further extract features such as average shortest path, closeness centralities, and clustering coefficients. Then, we condense the PLI connectivity matrix into a PLI connectivity vector by removing redundant information, to reduce dimensionality. Finally, the PLI connectivity vector and graph features are employed in classification. Fig. 3 shows the PLI connectivity matrices for the first five seconds or 640 samples of the low and high mental workload EEG data.

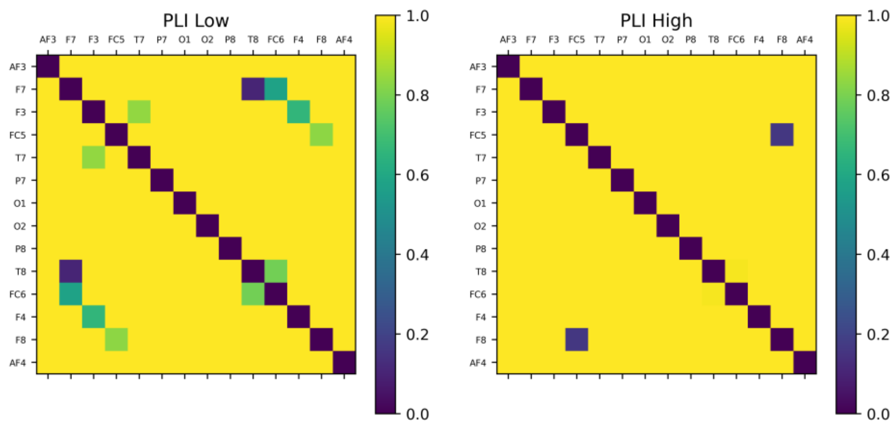

Fig. 3. PLI Connectivity Matrices for Low vs. High Mental Workload

2.3 Coherence

Coherence is the other feature we used to classify between two levels of mental workload. Coherence, also called magnitude-squared coherence, is a statistic used to examine the relationship between two signals at a specific frequency. Coherence starts with the idea of cross spectral density, which is a spectral analysis comparing two signals [2]. Cross spectral density is defined as the Fourier Transform of the cross-correlation function of signal X and signal Y, which is just a function that defines the relationship between two signals [2]. Auto spectral density is the same thing except we take the Fourier Transform of the cross-correlation function of signal X and itself, also called the auto-correlation function. Finally, coherence is defined as the following, where *f = frequency*, $C_{xy}(f)$ *= coherence between signal X and Y*, $G_{xy}(f)$ *= Cross spectral density between signal X and Y*, $G_{xx}(f)$ *= Auto spectral density of signal X*, and $G_{yy}(f)$ *= Auto spectral density of signal Y*:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} \tag{2}$$

Note that coherence, $C_{xy}(f)$, itself is actually function that takes in a specific frequency, *f*, and its output value is the actual statistic we are interested in. The resulting coherence statistic for a given frequency is a value between 0 and 1, where 0 indicates signal X and signal Y are completely unrelated, while a value greater than 0 indicates that either signal X and signal Y are dependent, the cross-correlation function between X and Y is non-linear, or there is noise in the measurements.

For feature extraction, we calculated the average coherence over 5 different frequency ranges, between all 14 sensors. These frequency ranges include 4-8 Hz (theta), 8-11 Hz (low-alpha), 11-13 Hz (high-alpha), 13-30 Hz (beta), and 30-40 Hz (gamma). The result is 5 different coherence connectivity matrices, which similarly to PLI, we used as graph adjacency matrices to further extract features. Then, we condensed the coherence connectivity matrices into connectivity vectors by removing redundant information and employed them along with the computed graph features in classification. Fig. 4 shows the coherence connectivity matrices for the first five seconds or 640 samples of the low and high mental workload EEG data, for each of the frequency ranges.

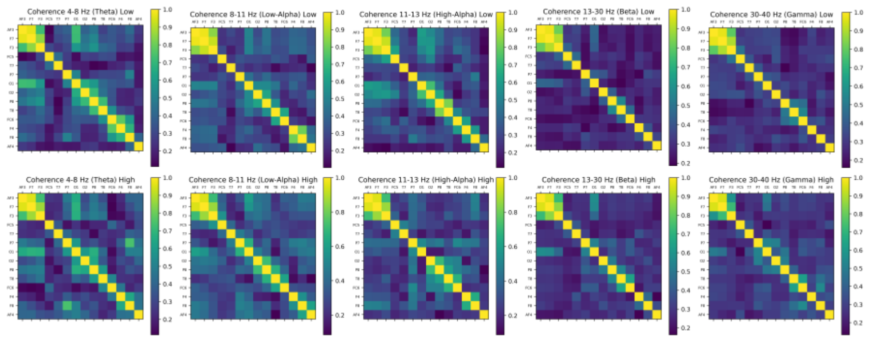

Fig. 4. Coherence Connectivity Matrices for Low vs. High Mental Workload

2.4 Classification

With our PLI and coherence features extracted from the EEG, we are finally able to begin classification on whether a specific sample of EEG data is associated with high or low mental workload. We begin by reading in the features associated with high and low mental workload into their own respective data frames. The data frames consist of a row for every 5 second interval, and a column for every feature value coming from PLI and/or coherence. Then, we create a new column in both data frames called "label", and we assign labels of 0 and 1 for low and high mental workload respectively. Afterward, we combine the low and high mental workload data frames and shuffle them, making them ready for classification.

We chose to test a variety of different ML models for the sake of analysis, namely support vector machine (SVM) and logistic regression (LOG) for their simplicity, random forest (RFC) for its reliability, and multi-layer perceptron (MLP) for its ability to deal with complex problems. For SVM, we used a linear kernel with C=1.0. For logistic regression, we let maximum iterations be 5000. For random forest, we tested it with n = 5, 10, 20, 30, 40, and 50 decision tree estimators. For multi-layer perceptron, we similarly tested it with n = 10, 20, 30, 40, 50, 100, and 250 hidden layer nodes for a single hidden layer, and 1000 maximum epochs. For scoring metrics, we collect information on the average fitting/scoring time, accuracy, f1 score, and log loss from ten-fold cross validation for both the training and the testing set. After classification, the results were stored into file for further analysis.

Note that since the main goal was to determine whether PLI or coherence features are more effective, we performed classification in three separate instances: one with only PLI features, one with only coherence features, and one with both PLI and coherence features.

3 Results

The results show the difference in efficacy of PLI and coherence features for different ML algorithms. All plots were generated with Python's Matplotlib library [3].

3.1 PLI Features Only

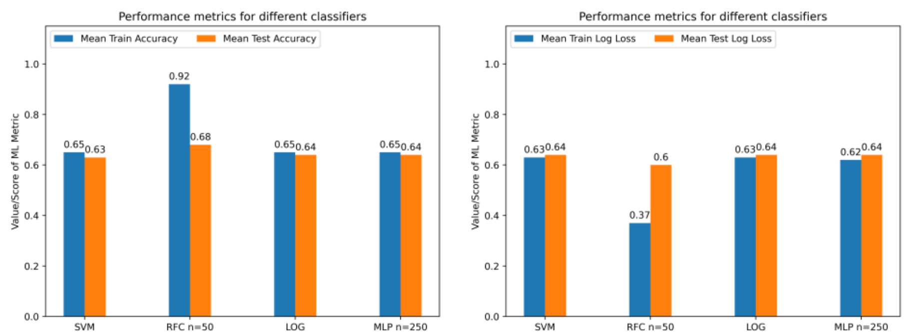

Fig. 5. Accuracy (Left) and Loss (Right) for PLI Features Only

Fig. 5 shows the average training and testing accuracy of different classifiers for PLI features only. The average testing accuracy and loss was pretty consistent for different ML models, being around 65% and 62% respectively. Notably, the ML models did not perform very well with PLI features alone.

3.2 Coherence Features Only

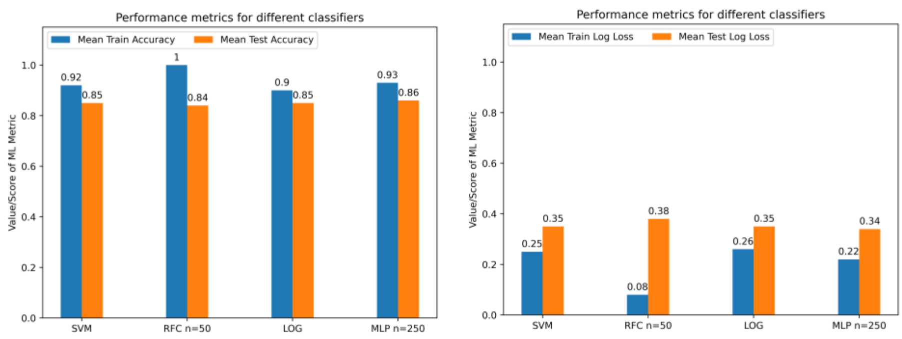

Fig. 6. Accuracy (Left) and Loss (Right) for Coherence Features Only

Fig. 6 shows the average training and testing accuracy of different classifiers for Coherence features only. The average testing accuracy and loss was pretty consistent for different ML models, being around 85% and 35% respectively. The ML models seemed to perform much better with the coherence features alone compared to just the PLI features.

3.3 PLI and Coherence Features

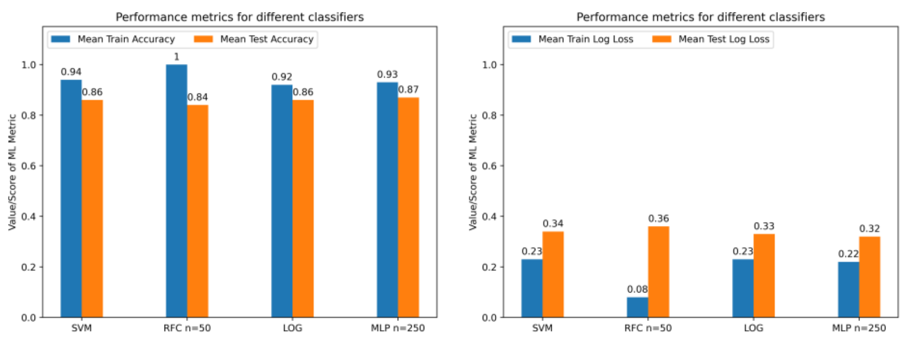

Fig. 7. Accuracy (Left) and Loss (Right) for PLI and Coherence Features

Fig. 7 shows the average training and testing accuracy of different classifiers for both PLI and Coherence features. The average testing accuracy and loss was around 86% and 35% respectively. The ML models seemed to perform just slightly better with both coherence and PLI features compared to just coherence features, but this difference is probably negligible.

4 Discussion

Based on the results, using the coherence features were much more effective in classification between two levels of mental workload. Fig. 6 illustrates that using coherence features alone yielded an “acceptable” accuracy of around 85%, while Fig. 5 illustrates that only using PLI features gave an accuracy of around 65%. This result is consistent across the different ML algorithms we used, namely SVM, RFC, LOG, and MLP.

Some limitations of the methodology include only testing four different ML algorithms, as we could have gotten different results with other untested ML algorithms. In addition, we did not extract frequency ranges for computing the PLI features like we did for the coherence features, which may help to explain the huge difference between the effectiveness of the PLI and coherence features. In addition, we performed offline classification for this project, selecting 5 second intervals to get as close to real-time classification as possible. We also only used three methods of graph connectivity, which were average shortest path, closeness centralities, and clustering coefficients.

In the future, we will examine if we can get better results with more complex ML algorithms such as convolutional, transformer, or graph neural networks. In addition, we expect to get much better results and performance by extracting PLI features for each of the 5 frequency ranges, like we did for coherence. We would like to explore additional connectivity methods, such as imaginary coherence and weighted PLI, investigating their effectivity and experimenting to maximize the accuracy of the ML model. We would also like to work with different datasets to verify that our models and methods of connectivity work in general. It would also be insightful to experiment with different length time intervals, such as 10 seconds, or 2.5 seconds to get “closer” to real-time passive BCI. We also want to apply more rigorous saliency methods to determine which specific features contributed more to the performance of the ML algorithms.

References

1. Wei Lun Lim, Olga Sourina, Lipo Wang. (2018). STEW: Simultaneous Task EEG Workload Dataset. *IEEE Dataport*. https://dx.doi.org/10.21227/44r8-ya50
2. *Cross Power Spectral Density Spectrum for noise modelling and filter design*. Cross Power Spectral Density Spectrum for Noise Modelling and Filter Design. (2023, August 11). https://resources.system-analysis.cadence.com/blog/msa2020-cross-power-spectral-density-spectrum-for-noise-modelling-and-filter-design
3. Hunter, J. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
4. Alimardani, M., & Hiraki, K. (2020). Passive Brain-Computer Interfaces for Enhanced Human-Robot Interaction. In Frontiers in Robotics and AI (Vol. 7). Frontiers Media SA. https://doi.org/10.3389/frobt.2020.00125
5. *BCI guide*. EMOTIV. (2023, March 21). https://www.emotiv.com/bci-guide/
6. *Mental workload*. Mental Workload - an overview | ScienceDirect Topics. (n.d.). https://www.sciencedirect.com/topics/engineering/mental-workload

7. *Functional connectivity in the Brain: Research Project*. The Florey. (2023, April 23). https://florey.edu.au/research-project/functional-connectivity-in-the-brain/