Teymourlouei, A., Hu, M., Gentili, R., & Reggia, J. (2024, July). Functional connectivity methods for multi-class mental workload classification. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-4). IEEE. Link: https://ieeexplore.ieee.org/abstract/document/10782848.

Functional Connectivity Methods for Multi-Class Mental Workload Classification

Arya Teymourlouei*
Department of Computer Science
University of Maryland
College Park, USA
ateymo2@terpmail.umd.edu

Minsi Hu*
Department of Computer Science
University of Maryland
College Park, USA
mhu0315@umd.edu

Rodolphe Gentili
Department of Kinesiology
University of Maryland
College Park, USA
rodolphe@umd.edu

James Reggia
Department of Computer Science
University of Maryland
College Park, USA
reggia@umd.edu

***Abstract*—Recently, significant attention has been drawn to the ability of network-based features to classify EEG signals reflecting varying levels of mental workload. Such features are based on methods of functional connectivity (FC), which quantify the statistical relationship between EEG electrode potentials. Here, we compare three FC-based feature extraction methods for the classification of mental workload from the Multi-Attribute Task Battery. The approaches used are weighted phase lag index (WPLI), imaginary coherence (IC), and layer entanglement (LE). WPLI and IC are popular methods for FC analysis. LE is a new approach which was introduced in recent literature. When classifying between three levels of workload, a support vector machine classifier achieved an 88% average (person-dependent) accuracy using all FC methods together, 89% using only the LE method, 67% with the IC method, and 61% with the WPLI method. When classifying between two levels of workload, these scores improve to 97%, 97%, 86%, and 81%, respectively. These results support and extend the findings of prior work and suggest that LE-based methods may enable accurate mental workload prediction which is suitable for passive brain-computer interfaces.**

***Keywords*—Classification, support vector machine, functional connectivity, electroencephalography (EEG) analysis, mental workload prediction**

I. INTRODUCTION

The rapid advancement of computer technology over the last three decades has enabled large-scale, data-driven research in a variety of fields. In neuroscience, this has led to an influx of massive, complex neural data sets which can be challenging to effectively analyze [1]. However, recently, techniques from network science have helped make great progress in interpreting such data [2]. One example is functional connectivity (FC), methods which can investigate either the statistical relationship between sources of neural activity (source space) or between sensor/electrode potentials (sensor space) [3]. FC methods create functional brain networks, which can be used to study the strength in connectivity between brain regions. One way to study FC is with an electroencephalography (EEG) device, which records neural activity from the scalp. EEG is useful for measuring the activity associated with mental workload [4], which is defined as the neural resources engaged to face task demand during performance [5, 6]. In this research, we investigate the ability of three different FC methods to accurately classify varying levels of mental workload from EEG signals.

In recent work [7], we demonstrated a novel approach to accurately predicting mental workload based on EEG signals collected from the Simultaneous EEG Task Workload Dataset [8]. Our feature extraction was based on transforming EEG time series to complex networks using the visibility graph (VG) algorithm. The VGs of ten EEG channels were created and assembled into a single multiplex temporal network (MTN). Each graph in the MTN is referred to as a layer. The nodes in each layer, corresponding to the time points from the time series, are identical (only edges differ). For features, we constructed matrices (which were denoted as C) based on the share of overlapping edges between individual layers in the MTN. This measurement is called the edge overlap, or layer entanglement (LE). LE quantifies the similarity of the structure of the layers in the MTN. Although LE is not yet a validated method for FC, the C matrices constructed in [7] were based on electrode sensor pairs, as is typically seen in sensor-space FC. We found that a support vector machine (SVM) classifier trained with features related to LE achieved a 97% mean testing accuracy and 0.07 mean testing loss when classifying between two levels of mental workload. However, this work had three limitations. First, we did not investigate how the method performed in multi-class mental workload prediction. This is an issue because in real-time applications, mental workload can fluctuate beyond two states. Also, in [7], we did not study how our method compares with any of the existing FC methods. Finally, we only tested the LE method on one EEG dataset (STEW).

Here we expand on our prior work by addressing the major limitations explained above. Specifically, a new EEG dataset was used to perform classification between three levels of mental workload (associated with three levels of a cognitive task). The machine learning (ML) algorithms were trained with three different FC-based feature extraction methods: weighted phase lag index (WPLI), imaginary coherence (IC), and the LE

Acknowledgment: This work was supported in part by the Maryland Space Grant Consortium.

*Authors contributed equally.

method described above. WPLI and IC are two of the most popular FC methods for EEG analysis [9, 10]. The performance of the ML algorithms is compared when using each of the methods above, in both ternary and binary classification.

II. Methodology

The methodology encompasses EEG data collection and preprocessing, FC-based feature extraction, and mental workload prediction (in binary & ternary scenarios). A visual illustration of the process is presented in Fig. 1.

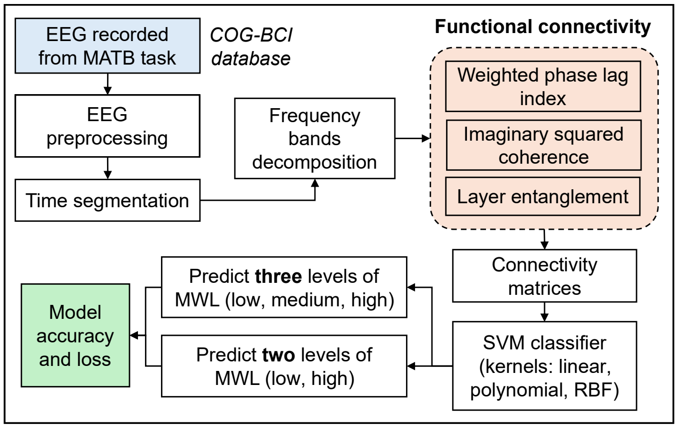

Fig. 1. Diagram of the methodology in this research. The acquired EEG data is preprocessed using conventional techinques. Then, FC-based measures are extracted and sent to an SVM classifier for prediction of mental workload.

A. EEG Dataset

A subset of the EEG data in the open-access COG-BCI database is used [11]. COG-BCI consists of EEG collected from four tasks: The Revised Multi-Attribute Task Battery (MATB-II) task, the N-Back task, the psychomotor vigilance task, and the Flanker task. In our computational pipeline, we will only consider the EEG from the MATB task. The National Aeronautics and Space Administration developed the MATB-II software as a means of evaluating mental workload dynamics and performance during the completion of simulated aircraft tasks. These tasks include monitoring system status, managing incoming communications, navigating a moving target, and managing fuel tank levels. A full description of these tasks is available in [12]. In [11], 29 participants completed the MATB under three levels of difficulty. In the easy condition, only two of the four subtasks were completed simultaneously. In the medium condition, this was increased to three subtasks. In the difficult condition, participants completed all four subtasks at the same time. Participants completed the tasks in fifteen minutes, split across three sessions. The EEG was recorded with a 64-electrode system using an ActiCap headcap (Brain Products, GmbH). To reduce the computational complexity of the feature extraction, only the 10 of these electrodes were used. These are F3, F4, C3, C4, P3, P4, T7, T8, O1, and O2.

B. EEG Preprocessing

Some steps were applied to clean the EEG data. First, a 1-50 Hz frequency impulse response filter was applied to each of the signals. Then, signals were re-referenced to the common average reference. Independent component analysis was not performed to assess the robustness of the classification when using data not completely free of artifacts, as this is the case during real-time classification. After cleaning, EEG signals were segmented into 10-second windows. The recorded sampling rate was 500 Hz, so, there were 5000 data samples for each time window. The data for the last ten seconds of each session were discarded to ensure the same data size for each participant. Five frequency ranges were extracted from the signals: theta (4-8 Hz), low-alpha (8-10 Hz), high-alpha (11-13 Hz), beta (13-30 Hz), and gamma (30-40 Hz). These ranges have been observed to reflect changes in mental workload in prior work [6, 13]. Features were computed separately for each frequency range.

C. Weighted Phase Lag Index

In this research, we compare the classification performance using three different feature extraction methods. Each of these are related to sensor-space FC: the statistical relationships between the electrode potentials from two EEG sensors. When computed for all sensor pairs, this allows for the formation of a connectivity matrix. The first FC method is the WPLI, defined in [9]. WPLI is less sensitive to volume conduction than the original phase lag index [9]. Let E[X] refer to the expected value of X, Im(ω) refer to the imaginary component of ω, and $C_{x,\,y}$ refer to the cross-spectral density of two signals x and y. The WPLI for two signals x and y is defined as:

$$\mathrm{WPLI}(x, y) = \frac{\left|\mathrm{E}\left[\mathrm{Im}\left(C_{x,\,y}\right)\right]\right|}{\mathrm{E}\left[\left|\mathrm{Im}\left(C_{x,\,y}\right)\right|\right]} \tag{1}$$

The WPLI was computed for all sensor pairs, so connectivity matrices with size (10, 10) were formed for each EEG segment.

D. Imaginary Coherence

The second method of assessing FC between EEG signals is the IC, defined in [10]. The IC is equivalent to the typical formulation of coherence, except that the imaginary component of the cross-spectral density is taken. Prior work has shown that the IC variant is less susceptible to volume conduction in EEG signals [10]. The IC for two signals x and y, is defined as:

$$\mathrm{IC}(x, y) = \frac{\mathrm{Im}\left(C_{x,\,y}\right)}{\sqrt{\left(C_{x,\,x}\right)\left(C_{y,\,y}\right)}} \tag{2}$$

where $C_{x,\,x}$ is the power-spectral density of x (this is the equivalent of taking the cross-spectral density between x and itself). Like the WPLI approach, the IC was computed for all sensor pairs, so again (10, 10) matrices were formed.

E. Layer Entanglement

In recent work, we demonstrated the effectiveness of a novel connectivity-based feature extraction method for mental workload classification [7]. This method was based on two steps. The first step was to transform each EEG time series into a graph by means of the VG algorithm. The VG algorithm is defined for all data points on the time series [14]. Consider two data points i and j, with associated values y_i and y_j, and an arbitrary point y_k. The data points i and j are represented as nodes in a graph, and are connected with a single edge if:

$$y_k < y_i + \left(y_j - y_i\right)\left(\frac{k - i}{j - i}\right) \ \forall\, k \in (i, j) \tag{3}$$

Edges in the VG are directed and unweighted. The direction of an edge is determined by the y-values of the associated data points, where the source node has the larger y-value. The ts2vg [15] and NetworkX [16] libraries were used for VG assembly.

The second step of the method in [7] was to construct VGs for each of the channels and assemble them into a MTN. Then, the similarity of layers in the MTN were analyzed by computing the share of overlapping edges between them (called the edge overlap, or LE). This is done for all sensor pairs. The definition of the C matrix follows from [17]. Let $\boldsymbol{V}$ equal the set of all nodes in the MTN, and $\boldsymbol{E} = \boldsymbol{V} \times \boldsymbol{V}$ equal the set of all possible edges in the MTN (for any layer). Then, the C matrix is defined as:

$$C = (c_{i,j}) \tag{4}$$

where $(c_{i,j})$ denotes an entry in the matrix for layers i and j. Now we define the matrix elements. Let $p_{i,j}$ refer to the number of edges that two layers, i and j, have in common. Let $p_{i,i}$ refer to the number of edges in layer i. Then, the non-diagonal elements are defined as:

$$(c_{i,j}) = \frac{p_{i,j}}{p_{i,i}} \tag{5}$$

Let $|\boldsymbol{E}|$ refer to the maximum possible number of edges in any given layer (always the same, since nodes in each layer are identical). Then, the diagonal elements are defined as:

$$(c_{i,i}) = \frac{p_{i,i}}{|\boldsymbol{E}|} \tag{6}$$

The C matrix is computed for each EEG segment (with size (10,10), as before) and forms the final of the three FC methods which are used in the feature extraction process.

F. Feature Selection

Feature selection for the ML algorithms was conducted by means of a statistical analysis. Features were first evaluated for normality using a Shapiro-Wilk test. Those which followed a normal distribution were tested for statistical significance with either a paired t-test (for binary classification between low and high mental workload) or a one-way ANOVA test (for ternary classification between three levels of mental workload). If the features did not follow a normal distribution, significance was assessed with either a Wilcoxon test (two levels of workload) or Friedman test (three levels of workload). Features which showed a statistically significant difference ($p < 0.05$) between the levels of mental workload were selected as input to the ML algorithm.

G. Classification

The features which pass the selection process were used as inputs to a support-vector machine (SVM) classifier. In prior work, SVM was shown to classify EEG signals associated with mental workload with over 97% accuracy, which was superior to several other classifier methods tested [7]. We tested the performance of the SVM algorithm using three variations of the kernel function: linear (C = 1.0), polynomial (C = 0.92, degree=2), and radial-based (RBF) (C = 1.0). Here, C refers to the regularization hyperparameter, unrelated to the C matrix defined in section D above. The gamma hyperparameter was computed as $1/(n \times \text{variance}(X))$, where n = number of features used, and X = the array of all features. The SVM was implemented in Scikit-Learn [18]. Classification was performed separately for each subject's data. The classification performance was assessed with two metrics: accuracy and loss. Accuracy refers to the fraction of correct predictions. The loss score quantifies the magnitude of error made by the classification model. The logistic loss function is used. Scoring was performed with a ten-fold cross-validation procedure.

III. Results

The results demonstrate the ability of the ML algorithms to discriminate between both three (section A) and two (section B) levels of mental workload using FC matrices from EEG signals. All plots in this section were generated with Python's Matplotlib [19] and Seaborn [20] plotting libraries.

A. Ternary Classification Performance

In this section, we examine the classification performance for an SVM tasked with predicting three levels of mental workload corresponding to three levels of difficulty in the MATB task. Fig. 2 shows the mean accuracy and loss scores (for testing data) using four methods for feature extraction: LE (the C matrices only), IC matrices only, WPLI matrices only, or matrices from all three methods combined (ALL). In the plot, vertical lines indicate standard error. Scores are reported for three variants of the SVM kernel: RBF (rbf-svm), polynomial (poly-svm), and linear (lin-svm).

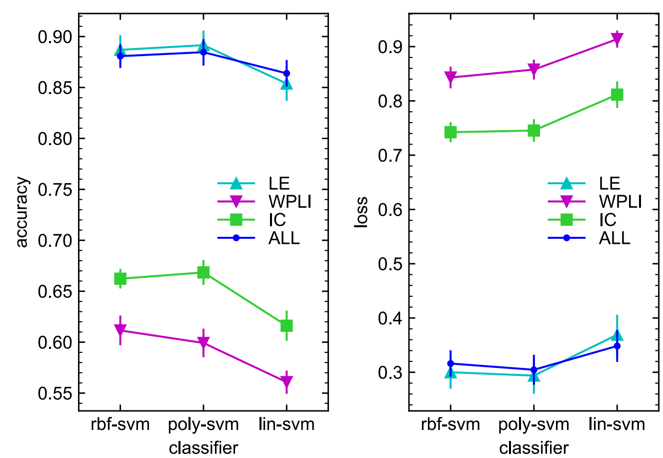

Fig. 2. *Left plot*: The mean testing accuracy scores and standard error for three SVM-based classifiers (variations of the kernel: radial-based function, polynomial, linear) to predict **three** levels of mental workload. Scores are computed for each participant (n=29). *Right plot*: Associated loss scores.

The highest mean accuracy in Fig. 2 was 89% with the polynomial SVM using the LE method. The same score was achieved when using all FC methods together. The lowest average loss was 0.30, with the same configuration. The highest mean accuracy achieved by the IC method was 67% (with polynomial SVM). For the WPLI method, this score was 61% (with RBF SVM).

B. Binary Classification Performance

Here, we investigate the classification performance of SVM to perform binary classification (between two levels of mental workload, corresponding to the easy and hard levels of difficulty in the MATB task). In Fig. 3, we present the mean accuracy and loss scores for three variants of SVM, using the same four feature extraction methods as section A. The only difference is that the classifier only must discriminate between two (as opposed to three) levels of mental workload. The style of the plot is the same as Fig. 2.

Fig. 3 shows that the highest average accuracy achieved was 97% with the polynomial SVM using the LE method (or all methods combined) for features. The SVM trained with RBF kernel also achieved 97% mean accuracy using either LE method or all FC methods combined. The lowest average loss score achieved was 0.1, reached both by the polynomial and RBF kernels. The best average accuracy with IC features was 86% (RBF kernel), and 81% (RBF kernel) for WPLI features.

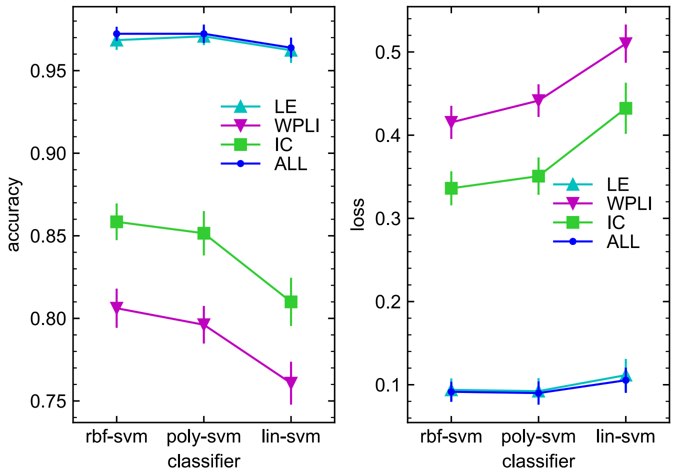

Fig. 3. *Left plot*: The mean testing accuracy scores and standard error for three SVM-based classifiers (variations of the kernel: linear, polynomial, radial-based function) to predict **two** levels of mental workload. Scores are computed for each participant (n=29). *Right plot*: Associated loss scores.

IV. Discussion

In this work, we investigated the effectiveness of three connectivity methods for classifying mental workload from EEG signals. We studied three levels of workload corresponding to three levels of difficulty in the MATB task. When tasked with predicting between all three levels, the SVM classifier trained with only the LE features achieved an average 89% accuracy and 0.3 loss across all subjects. This was an improvement over using only IC matrices (67% accuracy) or using only WPLI matrices (61% accuracy). In the case of binary classification (only low and high workload), the SVM trained with only LE matrices achieved an average 97% accuracy and 0.1 loss across all subjects. Again, this performance exceeded that of the IC (86% accuracy) and WPLI (81% accuracy) methods.

The classification performance achieved in this research is like our prior work which used features related to VGs and MTNs [7, 13]. However, in this work, we addressed a limitation of [7], that the classification was done between only two levels of mental workload. Also, here we expanded on prior research by comparing our approach with some of the more popular connectivity methods [9, 10]. Finally, in this work, we tested the ability of the LE method to generalize to a new EEG dataset (beyond the STEW data which was used in [7] and [13]).

The results of this work are relevant to many bioengineering applications. For example, the results of FC analysis can be used to study the cognitive-motor mechanisms associated with increases in mental demand during task completion [21]. Further, this work can help enable real-time decoding models in passive brain-computer interfaces [22]. To this end, we suggest future work to explore the capability of FC-based classification models to detect mental workload in real-time context. Other FC methods should also be compared, such as amplitude envelope correction [23]. Finally, one major remaining limitation of this work is the neurophysiological interpretation of the VG & LE methods. We strongly recommend investigations (e.g., with simulated EEG) that can help characterize these approaches.

References

[1] Frégnac, Y. (2017). Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain? *Science*, *358*(6362), 470-477.

[2] Bassett, D., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353-364.

[3] Bastos, A., & Schoffelen, J. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience*, *9*, 175.

[4] Wickens, C. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449-455.

[5] Young, M., Brookhuis, K., Wickens, C., & Hancock, P. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17.

[6] Jaquess, K., Lo, L., Oh, H., Lu, C., Ginsberg, A., Tan, Y., ... & Gentili, R. (2018). Changes in mental workload and motor performance throughout multiple practice sessions under various levels of task difficulty. *Neuroscience*, *393*, 305-318.

[7] Teymourlouei, A., Stone, J., Gentili, R., & Reggia, J. (2023). Multiplex temporal networks for rapid mental workload classification. In *International Conference on Brain Informatics* (pp. 203-214).

[8] Lim, W., Sourina, O., & Wang, L. (2018). STEW: Simultaneous task EEG workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2106-2114.

[9] Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., & Pennartz, C. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, *55*(4), 1548-1565.

[10] Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*, 2292-2307.

[11] Hinss, M., Jahanpour, E., Somon, B., Pluchon, L., Dehais, F., & Roy, R. (2023). Open multi-session and multi-task EEG cognitive dataset for passive brain-computer interface applications. *Scientific Data*, *10*(1), 85.

[12] Santiago-Espada, Y., Myer, R., Latorella, K., & Comstock Jr, J. (2011). The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide (No. L-20031)).

[13] Teymourlouei, A., Gentili, R., & Reggia, J. (2023). Decoding EEG signals with visibility graphs to predict varying levels of mental workload. In *57th Annual Conference on Information Sciences and Systems* (pp. 1-6).

[14] Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuno, J. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, *105*(13), 4972-4975.

[15] Bergillos, C. (2020). A study of visibility graphs for time series representations. Universitat Politècnica de Catalunya.

[16] Hagberg, A., Swart, P., Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Lab.

[17] Škrlj, B., & Renoust, B. (2020). Layer entanglement in multiplex, temporal multiplex, and coupled multilayer networks. *Applied Network Science*, *5*(1), 1-34.

[18] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *J. of Machine Learning Res.*, *12*, 2825-30.

[19] Hunter, J. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90-95.

[20] Waskom, M. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

[21] Shaw, E., Rietschel, J., Shuggi, I., Xu, Y., Chen, S., Miller, M., ... & Gentili, R. (2019). Cerebral cortical networking for mental workload assessment under various demands during dual-task walking. *Experimental Brain Research*, *237*, 2279-2295

[22] Yadav, H., & Maini, S. (2023). Electroencephalogram based brain-computer interface: Applications, challenges, and opportunities. *Multimedia Tools and Applications*, 1-45.

[23] Schoonhoven, D., Briels, C., Hillebrand, A., Scheltens, P., Stam, C., & Gouw, A. (2022). Sensitive and reproducible MEG resting-state metrics of functional connectivity in Alzheimer's disease. *Alzheimer's Research & Therapy*, *14*(1), 38.