Classifying Subjects with PFC Lesions from Healthy Controls during Working Memory Encoding via Graph Convolutional Networks

Sai Sanjay Balaji
Dept. Electrical & Computer Engineering
University of Minnesota, Minneapolis, USA
balaj037@umn.edu

Keshab K. Parhi
Dept. Electrical & Computer Engineering
University of Minnesota, Minneapolis, USA
parhi@umn.edu

***Abstract*—This paper describes a group-level classification of 14 patients with prefrontal cortex (pFC) lesions from 20 healthy controls using multi-layer graph convolutional networks (GCN) with features inferred from the scalp EEG recorded from the encoding phase of working memory (WM) trials. We first construct undirected and directed graphs to represent the WM encoding for each trial for each subject using *distance correlation*-based functional connectivity measures and *differential directed information*-based effective connectivity measures, respectively. Centrality measures of betweenness centrality, eigenvector centrality, and closeness centrality are inferred for each of the 64 channels from the brain connectivity. Along with the three centrality measures, each graph uses the relative band powers in the five frequency bands - delta, theta, alpha, beta, and gamma- as node features. The summarized graph representation is learned using two layers of GCN followed by mean pooling, and fully connected layers are used for classification. The final class label for a subject is decided using majority voting based on the results from all the subject's trials. The GCN-based model can correctly classify 28 of the 34 subjects (82.35% accuracy) with undirected edges represented by functional connectivity measure of distance correlation and classify all 34 subjects (100% accuracy) with directed edges characterized by effective connectivity measure of differential directed information.**

***Index Terms*—brain connectivity, graph convolutional networks (GCN), prefrontal cortex (pFC), working memory task**

I. Introduction

Working memory (WM) in humans is the brain system that provides the capability of actively storing information over short periods, which is essential to perform any complex cognitive task [1]. WM is often impaired by neurological disorders such as Parkinson's disease (PD), as evidenced by successful diagnoses using standardized tests for WM evaluation [2]. However, it is possible to alter WM to treat neurological disorders. For example, modification of WM through therapy has been shown to treat anxiety symptoms and post-traumatic stress disorder (PTSD) [3]. Thus, understanding the neural pathways responsible for WM can ease the modification process, treat the symptoms of severe neurological disorders, and provide insight into human cognition.

The prefrontal cortex (pFC) of the frontal lobe, which has been instrumental in complex cognitive behavior, personality expression, decision-making, and moderating social behavior [4], also plays a vital role in the WM process. Brain imaging using MRI revealed a strong correlation between WM load and the activity of the pFC region inferred from positive, statistically significant pixel-wise signal differences induced by the stimulus [5]. Despite multiple studies corroborating the essential role of pFC in the WM process [6], [7], it is non-intuitive to observe patients being able to complete WM tasks despite suffering from severe tissue damage (lesions) to the pFC cortical region. This discovery hints at the possibility of alternate neural pathways for WM that do not heavily rely on pFC. Additional support for this hypothesis comes from the study's results in [8] that identified pFC-independent frontoparietal neural pathways for WM, and pFC is not always required for a successful WM process.

Traditional machine learning (ML) and deep learning (DL) algorithms have recently become prevalent in analyzing datasets of the brain's electrophysiology and have shown robust predictions with high accuracy [9]. However, they fail to provide the mapping of brain connectivity by locating Regions of Interest (ROIs) [10]. The fundamental network-like structure of the brain is often ignored in these models. A graph representation of the brain is thus better suited to explain the interactions by considering its physiological configuration. The graph edges can represent the two types of commonly used brain connectivity measures for statistical analysis. *Functional connectivity* characterizes the correlation between the different brain regions, which identifies the cluster of regions activated during a cognitive task, and *Effective connectivity* describes the *directional* flow of information/signal between different brain regions during a cognitive task [11].

This paper discusses a graph convolutional network (GCN)-based classifier to distinguish healthy controls from patients with lesioned pFC from scalp EEG recorded while performing WM trials. A graph data type represents the functional or directed functional connectivity as edge features, and node features represent each channel's centrality measures and band powers. These models can better characterize the variations in memory encoding of the two classes of subjects studied in our previous work [12]. The remainder of the paper is organized as follows. Section II provides a brief overview of WM trials and the data. Generation of graph and GCN classifier architecture are described in Section III. The classifier results and our

This paper was supported in part by the National Science Foundation under grant number CCF-1954749.

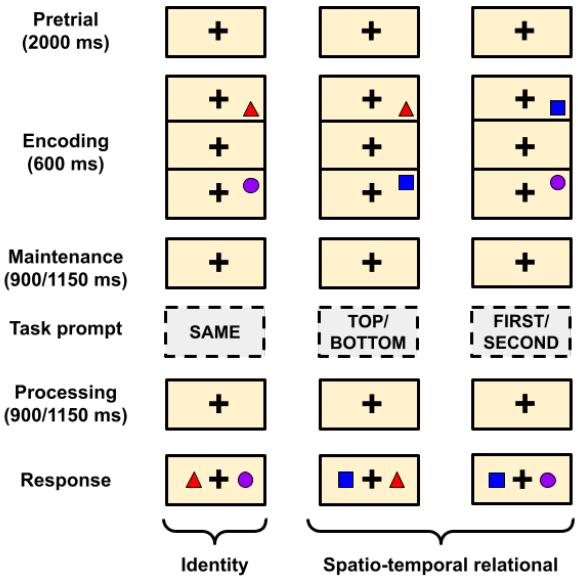

Fig. 1. The phases of a lateralized visuospatial WM task, adapted from [12].

inference are discussed in sections IV and V, respectively.

II. Overview of Task and Data

A. Working Memory (WM) task

The WM is assessed in two ways using *identity* and *spatio-temporal relation* tests [8]. In either case, the subjects are shown a pair of common shapes sequentially. For the identity test, a task prompt shows a pair of shapes, and the subjects are asked to identify whether the shown pair is the same as what they had just studied. For the relation test, the spatial or temporal aspects are assessed by prompting the subjects to indicate the shape observed in the top/bottom or first/second, respectively.

Each WM trial can be divided into five phases, as illustrated in Fig. 1. Central fixation is shown to record the resting state EEG during the 2 s *pretrial* phase. After this, subjects are shown two common shapes in a top/bottom spatial orientation for 200 ms each sequentially with a 200 ms break in between, marking the *encoding* phase. A 900 ms or 1150 ms *maintenance* interval follows the encoding phase, where the subjects actively hold the information shown during the encoding phase. This is followed by the *active processing* stage, where a text prompt appears for the same duration as the maintenance phase. Finally, the subjects indicated their *response*.

B. Data

The data set consists of scalp EEG recordings from 20 healthy control subjects and 14 patients with unilateral pFC lesions performing lateralized visuospatial working memory tasks [13]. All participants gave informed written consent following the University of California, Berkeley, Institutional Review Board, or the Regional Committee for Medical Research Ethics, Region South, per the Declaration of Helsinki.

TABLE I
Description of graph edge and node features for the three types of graphs generated for classification

S.No	Graph type	Edge features	Node features
1.	Undirected	Distance correlation (DC) between each channel pair during WM encoding **No. of edges** $=\binom{64}{2}$ = 2016	None (all set to 1 to infer the GCN performance using the weighted graph alone)
2.	Undirected	Distance correlation (DC) between each channel pair during WM encoding **No. of edges** $=\binom{64}{2}$ = 2016	Relative power in the five frequency bands and three centrality measures inferred from DC-edge-connected graph **No. of node features** = 8
3.	Directed	Differential DI (the change in DI value during WM encoding from the baseline) between each channel pair during WM encoding **No. of edges** = $2 * \binom{64}{2}$ = 4032	Relative power in the five frequency bands and three centrality measures inferred from differential DI-edge-connected graph **No. of node features** = 8

Using a 64 + 8 channel BioSemi ActiveTwo amplifier with Ag-AgCl pin-type active electrodes mounted on an elastic cap, the scalp EEG was recorded with a 1024 Hz sampling frequency. The participants completed 120-240 trials of working memory tasks each, with each trial having an equal probability of being an identity or relation type. The EEG signals from the 64 channels were recorded during the five phases. The raw signals were preprocessed for noise removal and normalizing all lesions to the left hemisphere using spatial transformation. The recordings for each trial were then segregated into three multivariate time-series corresponding to the pretrial phase, encoding and maintenance phase, and active processing phase.

III. Experimental Setup

A. Graph Generation

The preprocessed EEG signals are transformed into graphs with 64 nodes representing 64 channels and edges representing the adjacency matrix corresponding to either functional or directed functional connectivity (effective connectivity), as described in Table I. The following features are used for generating graphs from the EEG recordings:

- *Adjacency matrix*: We compare the classification performance of two different brain connectivity parameters to represent the adjacency matrix. The functional connectivity is expressed using the statistical *distance correlation* (DC) that measures the linear and nonlinear association between two random variables, i.e., EEG recordings from two channels [14]. Directed information (DI) [15], which statistically provides the degree of causation from the observed EEG recordings, is used to characterize the effective connectivity. The DI estimator was adapted from [16]. The superiority of DI-based connectivity features for subject-wise classification of the memory encoding phase from baseline is demonstrated in [12] using the same data used in this study. To combat inter-subject variances that stem from differences in subjects' physiology, we represent the connectivity features as the absolute change from the pretrial baseline to the encoding phase.

- *Node features*: A combination of Centrality features and relative power spectral densities (PSD) are used as the node features. *Betweenness centrality* (BC) is a topological feature that measures the fraction of shortest paths passing through a node [17] in a graph. Nodes with a high betweenness centrality are interpreted as *gatekeepers* as cutting off the edges to these nodes can split the graph into clusters, severing information flow across the network. Thus, BC is a reliable measure of a node's influence over information flow within a graph and is used to identify significant regions during cognitive tasks [18]. In addition to BC, we also included *eigenvector centrality* and *closeness centrality* to learn additional information on a node's influence. Spectral power is another proven feature used in the literature for classification tasks involving electrophysiological datasets of the brain, such as seizure prediction [19], and automated schizophrenia screening [20]. For our analysis, we computed the relative PSD for each of the 64 channels in the five frequency bands - δ (1 - 4 Hz), θ (4 - 8 Hz), α (8 - 13 Hz), β (13 - 30 Hz), and γ (30 - 80 Hz). Thus, each node (channel) is represented using 8 features (3 centrality and 5 PSD measures).

B. Classifier Architecture

Fig. 2 shows our proposed architecture of the GCN-based classifier. The input is the graph represented by its adjacency matrix and node features. Feature representation from the generated graph is performed using two layers of GCN, followed by mean pooling. Each layer consists of 64 nodes corresponding to the 64 channels with rectified linear unit (ReLU) activation function. To improve generalizability and avoid overfitting the training data, we employ dropout regularization for each layer with a probability of retention, p = 0.5. Two dense layers follow the GCN layers with 32 and 16 units, respectively, and both use ReLU activation. Finally, the classifier layer is implemented using a fully connected layer with two units and sigmoid activation. Other hyper-parameters, such as batch size, learning rate, number of epochs, and optimization algorithm, are tuned to obtain the best-performing model. We also employ a callback feature to reduce the learning rate by monitoring the validation loss.

We employ the leave-one-subject-out cross-validation (LOOCV) technique, with the trials corresponding to one subject held out for testing and the remainder used for training and recursively repeating the process until all subjects are tested. The training data is further segregated into independent training and validation sets in the ratio 7:3. In each case, a single label is assigned to the test subject using $\frac{2}{3}$ majority voting from the predicted labels of all the trials of the test subject. A majority failure is also considered a misclassification when calculating the overall performance.

IV. Results

After an exhaustive tuning of hyper-parameters, the best-performing model is chosen. The GCN model was trained for 200 epochs with a training batch size of 16. The Adam optimizer with an initial learning rate of 10^{-3} and binary cross-entropy loss function based on training accuracy was used for this purpose. During training, the callback function monitored the validation loss and reduced the learning rate if the loss plateaued. We evaluated the overall classifier performance as the percentage of 34 subjects identified correctly after majority voting with class labels "0" and "1" representing control subjects and patients with pFC lesions, respectively.

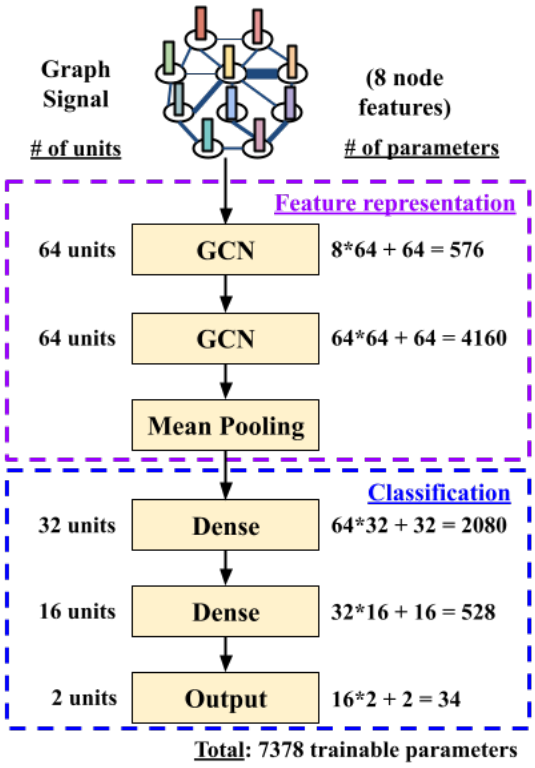

Fig. 2. Architecture of the GCN-based classifier showing the number of parameters for each layer: i) Two layers of GCN followed by mean pooling to learn the graph representation, ii) Two fully connected layers for classification.

We first evaluated the performance with DC features as edge weights. The classifier performance was compared with and without the node features, using the adjacency matrix as edge weights in both cases. After optimal hyper-parameter tuning, the GCN-classifier that utilized edge weights alone with no nodal information identified the class label of 6 of 14 subjects with pFC lesions (sensitivity of 42.86%) and 14 of the 20 control subjects (70% specificity), which gave an accuracy of 58.82%. The performance improved considerably with the addition of centrality and PSD measures. The GCN classifier with DC edge weights and the eight nodal features identified the class label of 11 subjects with pFC lesions (sensitivity of 78.57%) and 17 control subjects (85% specificity), resulting in an accuracy of 82.35%.

Using effective connectivity measures of *differential DI*, the change in DI feature during encoding from the pretrial baseline as edge weights instead of DC resulted in the best performance. The GCN classifier, in this case, identified the class label of all 34 subjects correctly (100% accuracy). Fig.

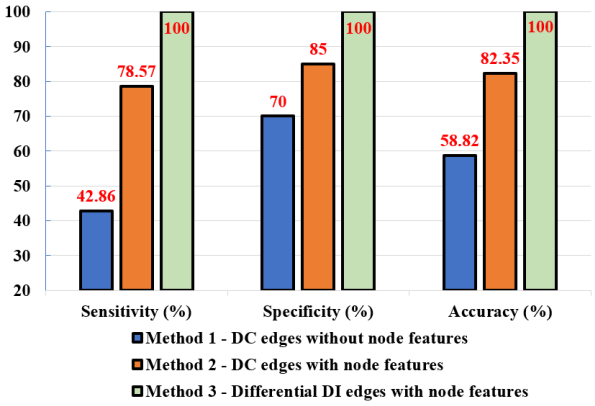

Fig. 3. Classification performance (% of subjects identified correctly using majority voting) with distance correlation (DC) and differential directed information (DI). The use of relative band power and centrality features as node features improves the classifier accuracy of the DC-edge-based classifier. Differential DI as edge features further improve the accuracy and lead to the identification of all 34 subjects.

3 summarizes the classification performance using the three methods discussed above.

V. Conclusion

This paper demonstrates the capability of GCN models to represent the cognitive features of WM encoding among two populations - healthy controls and patients with pFC lesions. The GCN model can perform group-level classification with high accuracy from functional and effective brain connectivity features, centrality measures inferred from the connectivity, and band powers from different channels. This indicates that a graph-based model of brain connectivity is well-suited for capturing relevant information across subjects, making them agnostic to inter-subject variances.

Amongst the two connectivity measures, the differential DI as directed edge weight results in the best performance, identifying the class label of all 34 subjects. Although DC can measure nonlinear interactions, DI-based effective connectivity measure is far superior in characterizing the nonlinear interaction in the brain, as indicated by its better performance (100%) when compared to that of DC (82.35%). Finally, we show that centrality measures and band powers play a crucial role in the classification process. The inclusion of these nodal features results in an increase in accuracy by 23.5%.

Though the number of parameters used by the GCN classifier is comparable to a simple neural network and much less than many standard deep neural networks, there is scope to simplify the architecture further. Future work will be directed towards deriving the two classes' graphical representation and identifying significant differences in the connectivity and nodal features. This can be extended to the decoding phase of WM tasks to identify the significant variations in regional dependence on the WM process among the two populations and can also be generalized to determine the deterioration in brain networks for subjects with memory disorders, such as dementia and Alzheimer's.

References

[1] Alan Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.

[2] Elizabeth A Kensinger, Deirdre K Shearer, Joseph J Locascio, John H Growdon, and Suzanne Corkin, "Working memory in mild alzheimer's disease and early parkinson's disease.," *Neuropsychology*, vol. 17, no. 2, pp. 230, 2003.

[3] Jackie Andrade, David Kavanagh, and Alan Baddeley, "Eye-movements and visual imagery: A working memory approach to the treatment of post-traumatic stress disorder," *British journal of clinical psychology*, vol. 36, no. 2, pp. 209–223, 1997.

[4] David R Euston, Aaron J Gruber, and Bruce L McNaughton, "The role of medial prefrontal cortex in memory and decision making," *Neuron*, vol. 76, no. 6, pp. 1057–1070, 2012.

[5] Jonathan D Cohen, Steven D Forman, Todd S Braver, BJ Casey, David Servan-Schreiber, and Douglas C Noll, "Activation of the prefrontal cortex in a nonspatial working memory task with functional MRI," *Human brain mapping*, vol. 1, no. 4, pp. 293–304, 1994.

[6] Todd S Braver, Jonathan D Cohen, Leigh E Nystrom, John Jonides, Edward E Smith, and Douglas C Noll, "A parametric study of prefrontal cortex involvement in human working memory," *Neuroimage*, vol. 5, no. 1, pp. 49–62, 1997.

[7] Shintaro Funahashi, "Prefrontal cortex and working memory processes," *Neuroscience*, vol. 139, no. 1, pp. 251–261, 2006.

[8] Elizabeth L Johnson, Callum D Dewar, Anne-Kristin Solbakk, Tor Endestad, Torstein R Meling, and Robert T Knight, "Bidirectional frontoparietal oscillatory systems support working memory," *Current Biology*, vol. 27, no. 12, pp. 1829–1835, 2017.

[9] Mohammad-Parsa Hosseini, Amin Hosseini, and Kiarash Ahi, "A review on machine learning for eeg signal processing in bioengineering," *IEEE reviews in biomedical engineering*, vol. 14, pp. 204–218, 2020.

[10] Andac Demir, Toshiaki Koike-Akino, Ye Wang, Masaki Haruna, and Deniz Erdogmus, "EEG-GNN: Graph Neural Networks for Classification of Electroencephalogram (EEG) Signals," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 1061–1067.

[11] Alard Roebroeck, Anil K Seth, and Pedro Valdes-Sosa, "Causal Time Series Analysis of Functional Magnetic Resonance Imaging Data," in *NIPS mini-symposium on causality in time series*. PMLR, 2011, pp. 65–94.

[12] Sai Sanjay Balaji, Sandeep Avvaru, and Keshab K Parhi, "Classification of Pretrial vs. Encoding stage for Working Memory Task among Subjects with pFC Lesions and Healthy Controls using Directed Information," in *2022 56th Asilomar Conference on Signals, Systems, and Computers, to appear*. IEEE, 2022.

[13] Elizabeth L. Johnson, "64-channel human scalp eeg from 14 unilateral pfc patients and 20 healthy controls performing a lateralized visuospatial working memory task.," 2017.

[14] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov, "Measuring and testing dependence by correlation of distances," *The annals of statistics*, vol. 35, no. 6, pp. 2769–2794, 2007.

[15] James Massey et al., "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, 1990, pp. 303–305.

[16] Sandeep Avvaru, Noam Peled, Nicole R Provenza, Alik S Widge, and Keshab K Parhi, "Region-Level Functional and Effective Network Analysis of Human Brain During Cognitive Task Engagement," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1651–1660, 2021.

[17] Douglas R White and Stephen P Borgatti, "Betweenness centrality measures for directed graphs," *Social networks*, vol. 16, no. 4, pp. 335–346, 1994.

[18] Mikail Rubinov and Olaf Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[19] Zisheng Zhang and Keshab K Parhi, "Low-complexity seizure prediction from ieeg/seeg using spectral power and ratios of spectral power," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 3, pp. 693–706, 2015.

[20] Tingting Xu, Massoud Stephane, and Keshab K Parhi, "Classification of single-trial meg during sentence processing for automated schizophrenia screening," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 363–366.