CLASSIFYING PATIENTS WITH PFC LESIONS FROM HEALTHY CONTROLS USING DIRECTED INFORMATION BASED EFFECTIVE BRAIN CONNECTIVITY MEASURED FROM THE ENCODING PHASE OF WORKING MEMORY TASK

Sai Sanjay Balaji, Graduate Student member, IEEE *and Keshab K. Parhi,* Fellow, IEEE

University of Minnesota, Dept. of Electrical & Computer Engineering
Minneapolis, USA

ABSTRACT

This paper describes a group-level analysis of 14 subjects with prefrontal cortex (pFC) lesions and 20 healthy controls performing multiple lateralized visuospatial working memory (WM) trials. Using effective brain connectivity measures inferred from directed information (DI) during memory encoding, we first show that DI features can correctly classify 18 control subjects and 11 subjects with pFC lesions, providing an overall accuracy of 85.3%. Second, we show that differential DI, the change in DI during the encoding phase from pretrial, can successfully overcome inter-subject variability and correctly identify the class of all 34 subjects (100% accuracy). These accuracy results are based on two-thirds majority thresholding among all trials. Finally, we use Welch's t-test to identify the crucial differences in the two classes' sub-networks responsible for memory encoding. While the *inflow* of information to the prefrontal region is significant among subjects with pFC lesions, the *outflow* from the prefrontal to the frontal and central regions is diminished compared to the control subjects. We further identify specific neural pathways that are exclusively activated for each group during the encoding phase.

***Index Terms*—** Directed information (DI), differential DI, Effective connectivity, Prefrontal cortex (pFC) lesion, Working memory

1. INTRODUCTION

Neuroscientists and psychologists have long been interested in identifying the underlying neural framework for human cognition. Any complex cognitive task requires a system for simultaneous storage and manipulation of the required information. Such a brain system is termed working memory (WM) [1]. The impairment of WM by neurological disorders such as Alzheimer's disease (AD) and Parkinson's disease (PD) has led to standardized tests to quantify WM and their use in diagnosing such disorders [2, 3]. Furthermore, the modification of WM through therapy has been shown to treat anxiety symptoms, and post-traumatic stress disorder (PTSD) [4].

Prior studies involving brain imaging have identified a linear relationship between the activity of the prefrontal cortex (pFC) and WM load, which proved an indispensable role of pFC in the memory process [5]. Electrophysiological recordings, such as EEG, inherently possess a higher temporal resolution than imaging, making them more suitable for describing the spatio-temporal dynamics of the brain networks during memory encoding and retrieval [6]. Analysis using EEG can provide a better understanding of the degree of dependence of WM on pFC activity. Newer research [7] proposes that pFC activity is not always necessary in WM tasks. However, further work is required to describe the consequences of damage to pFC tissues on memory encoding and the reason behind successful memory encoding and retrieval despite such impairment.

Studying the variations in brain connectivity during memory encoding can help address these unanswered questions. Brain connectivity is typically categorized into three levels: structural, functional, and effective [8]. *Structural connectivity* refers to the anatomical interconnections of neurons inside the brain that can be viewed using noninvasive imaging techniques. *Functional connectivity* provides the statistical correlation between various brain regions and identifies the cluster of active regions during a cognitive task. *Effective connectivity* quantifies the *directional* neural activity (flow of information) across different brain regions during a cognitive task. Functional and effective connectivities are inferred from the electrophysiological time-series simultaneously recorded from multiple electrodes.

Effective connectivity measures modeled using directed information (DI) [9] have shown excellent performance in classification tasks such as identification of seizure onset zone [10] and mental states pertinent to a cognitive task [11]. Our prior work demonstrated the superiority of DI-based effective connectivity in distinguishing the memory encoding phase from the pretrial baseline for a given subject [12].Similar to [13], this paper addresses group-level classification for distinguishing patients with pFC lesions from healthy controls. In [13], we demonstrated 100% classification accuracy using a two-layer graph convolution network (GCN) for feature representation from graph signals with *differential DI*, the variation of DI from the pretrial to the encoding phase, as edges and relative-band powers and centrality measures as node features. In contrast, in this paper, we address the classification using an SVM model with only the DI features from the encoding phase of WM trials and differential DI. From this, we infer that differential DI features are agnostic to the inter-subject variability seen in the group-level analysis. Furthermore, the SVM model is better suited for the analyzed data set due to its lower complexity in terms of the number of parameters and better explainability when compared to the GCN model.

2. EXPERIMENTAL SETUP

2.1. Overview of Working Memory task and data

We examined the human scalp EEG from subjects performing lateralized visuospatial working memory tasks [14]. The two classes of subjects studied are patients with unilateral pFC lesions ($n = 14$) and healthy control ($n = 20$). All participants provided written consent following the University of California, Berkeley, Institutional Review Board. The working memory is tested in two ways, as described in [7].

- *Identity test* - Subjects were shown a pair of shapes and then asked to identify whether a given pair of shapes was the same as what they had just observed.
- *Spatio-temporal relation test* - The subjects were initially shown a pair of shapes similar to identity tests. The spatial aspect was examined by cuing the subjects to indicate the shape observed in the top/bottom, and the temporal aspect was examined by cuing them to indicate the shape observed first/second.

There are five phases in each trial for a WM test. During the 2 s *pretrial* phase, central fixation is shown to record the resting state EEG. This is followed by the *encoding* phase, where subjects were shown two common shapes sequentially in a top/bottom spatial orientation for 200 ms each with a 200 ms break in between. Following a 900 ms or 1150 ms *maintenance* interval, a text prompt appears in the *active processing* stage that lasts for the same duration as the maintenance phase. Finally, the subjects indicated their *response*. Fig. 1 illustrates the five phases of a WM task.

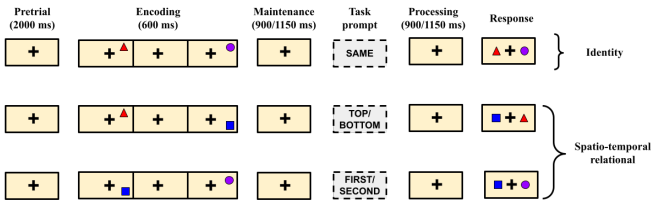


Fig. 1: The phases of a lateralized visuospatial WM task [12].

The 64 + 8 channel BioSemi ActiveTwo amplifier was used to record the scalp EEG [14] at 1024 samples/sec, with Ag-AgCl pin-type active electrodes mounted on an elastic cap. The electrophysiological signals from the 64 channels were recorded during each of the five phases of a WM trial. The preprocessing steps included spatial transformation for normalizing all lesions to the left hemisphere and noise removal. All subjects completed 120-240 trials, with each trial testing identity or relation with equal probability.

2.2. Feature extraction

2.2.1. Directed Information during the encoding phase

The information-theoretic measure of directed information (DI) [9] is used to characterize the brain's effective connectivity using the scalp EEG of the 64 channels recorded during the encoding phase of the WM trials. Let the time-series of scalp EEG recorded at channels *'x'* and *'y'* be denoted as $\mathbf{X}^N = [x_1, x_2, \ldots, x_N]$ and $\mathbf{Y}^N[y_1, y_2, \ldots, y_N]$, respectively, where N is the total sample length of the time-series. The DI from channel *'x'* to *'y'* is given by

$$\hat{I}(\mathbf{X}^N \rightarrow \mathbf{Y}^N) = \hat{h}(\mathbf{Y}^N) - \hat{h}(\mathbf{Y}^N||\mathbf{X}^N) \quad (1)$$

where $\hat{h}(\mathbf{Y}^N)$ and $\hat{h}(\mathbf{Y}^N||\mathbf{X}^N)$ represent the differential entropy of $\mathbf{Y}^N$ and the differential entropy of $\mathbf{Y}^N$ causally conditioned by $\mathbf{X}^N$, respectively [10]. The DI value $I(\mathbf{X}^N \rightarrow \mathbf{Y}^N)$ statistically provides an estimate of the degree to which $\mathbf{X}^N$ is relevant for causal inference on $\mathbf{Y}^N$ from the observed EEG recordings. The estimates of the two entropy values described in (1) are obtained using likelihood measures inferred from data-driven Kernel smoothing functions. The MATLAB implementation of the DI estimator was adapted from [11].

For the 64-channel scalp EEG recording, we have $2 \times \binom{64}{2} = 4032$ features representing the directional information from each channel to every other channel. Feature selection using the mutual information (MI) based minimal-redundancy-maximal-relevance (mRMR) framework [15] was employed to select up to 200 top directional connectivity as input features to the classifier. mRMR has shown to be a tractable option for dimensionality reduction in brain network classification problems for diagnosis of neural disorders, such as borderline personality disorder [16], schizophrenia [17], adolescent major depressive disorder [18], and attention-deficit hyperactivity disorder (ADHD) [19]. Other feature selection algorithms, such as MUSE [20], could also be used.

2.2.2. Differential DI

Group analyses involving multiple trials from various subjects suffer from inter-trial and inter-subject variances that stem from EEG recordings and the subject's physiology [21]. In our previous work [12], the high accuracy for subject-wise trial classification using DI features revealed that they are agnostic to inter-trial variance. However, the variations across different subjects that stem from physiological differences can obscure any similarity found across a group. Thus, we modeled another classifier that utilized the difference in DI value during the encoding phase from the pretrial baseline as its features.

2.3. Classifier

2.3.1. Classification using DI features

Fig. 2 illustrates the overall framework of the proposed preliminary architecture. We employ the leave-one-out (LOO) classification technique using linear SVM classifiers implemented in Python [22]. The nonlinear interactions across the different brain regions are already captured in the DI features. This, along with the moderately high dimensionality (up to 200), justifies using SVM classifiers with linear kernels. For each fold of the LOO classifier, the DI features obtained from the various trials of a subject are left out for testing, and the remainder was used for training.

Feature selection for each fold was performed separately based on the respective training data prior to classification. For each classifier, the model was recursively trained using the top 200 features ranked by the mRMR algorithm, and the best subset of features is identified as the one yielding the highest training accuracy. The classifier parameters were tuned using a comprehensive search. A single class label is assigned to the test subject based on $\frac{2}{3}$ majority voting after testing on all test subject's trials. We used the labels 0 and 1 to denote the control subjects and subjects with pFC lesions, respectively. SMOTE oversampling [23] prior to the classification was performed to address the class imbalance problem. The final performance of the classifier is determined based on the percentage of subjects identified correctly after the majority voting.

2.3.2. Classification using differential DI features

As such, the differential change in DI is smaller than the absolute value of DI during the encoding phase. Moreover, the change is relatively similar across the classes for many features. Only features with different mean values for the two classes with high statistical significance ($p < 0.05$) were considered for classification. Thus, only a small subset containing d of the total 4032 features will result in significant changes across the two groups, i.e. $d << 4032$. Fig. 3 depicts the overview of preprocessing steps used before classifying using relative DI features. The classification technique remained the same - Leave one subject out cross-validation using linear SVM. The feature dimensionality is indicated in red.

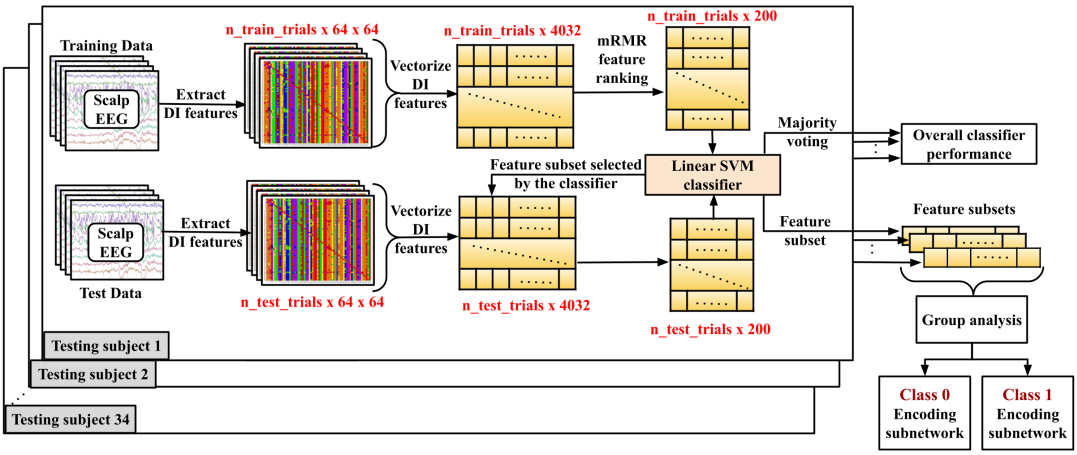


Fig. 2: Classification of healthy control and subjects with pFC lesions using DI features inferred from the encoding phase of WM tasks.

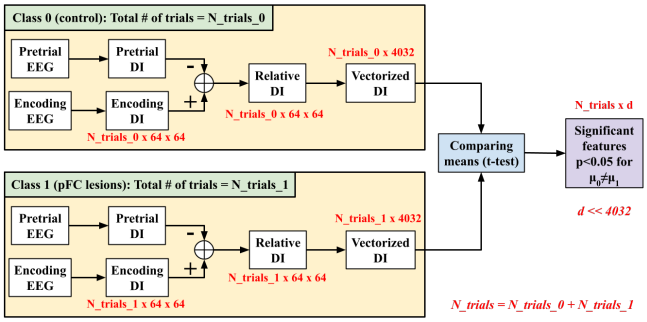


Fig. 3: Extraction and selection of differential DI features from training data for classifying control subjects vs. patients with pFC lesions (shown for one fold).

2.4. Group analysis

Sub-networks specific to the encoding phase are constructed from the subset of influential attributes obtained from feature selection that resulted in the best classifier performance. The mean DI and differential DI values are compared for the two classes to identify the statistically significant features that contrast the two populations. Following our previous work [12], the DI and differential DI features inferred from the 64 channels are then aggregated into eight regions - *anterio-frontal* (AF), *frontal* (F), *fronto-central* (FC), *central* (C), *centro-parietal* (CP), *parietal* (P), *parieto-occipital* (PO), and *temporal* (T).

3. RESULTS

3.1. Classification performance using DI features from the encoding phase

The best performance was observed from the model that used 135 of the top 200 features ranked using the mRMR algorithm. With an accuracy of 85.3%, the LOO classifier correctly identified 18 of the 20 healthy control (90.0% specificity) and 11 of the 14 subjects with pFC lesions (78.57% sensitivity). Given the smaller data size of 34 subjects, the inter-subject variability resulted in a marginally reduced performance.

3.2. Classification performance using differential DI features of the encoding phase from pretrial baseline

234 of the 4096 features revealed a statistically significant difference in the mean values of differential DI across the two classes. These features were used to classify the subjects with pFC lesions from healthy controls. Although the number of input features required is higher than the classifier that used absolute DI values (135), the differential DI features successfully identified the class of all 34 subjects using majority voting. This suggests that considering the change in DI value from baseline is an efficient method to combat inter-subject variability, which diminished the classifier's performance that only employed the DI features from the encoding phase. Table 1 summarizes the performance of the best models obtained from the two classification techniques used in this work.

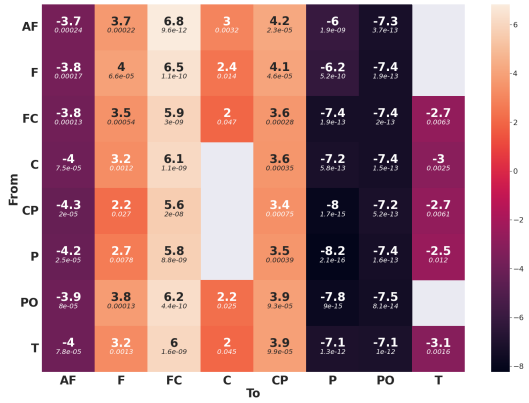
Table 1: Classification of subjects with pFC lesions (class 1) vs. healthy control (class 0) during the encoding phase of WM task

Features to the classifier	Specificity (%)	Sensitivity (%)	Accuracy (%)
DI (encoding phase)	90	78.6	85.3
Differential DI	100	100	100

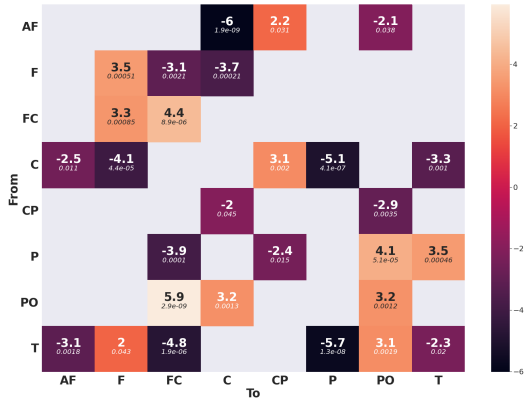
Note: Classification performance here is measured by the % of subjects identified correctly after $\frac{2}{3}$-majority voting and is not the trial accuracy of each fold.

3.3. Region-wise sub-network analysis

The two sets of features, DI values during the encoding phase and differential DI, were analyzed using Welch's t-test after performing region-wise aggregation. The heatmaps of the test results are shown in Fig. 4. The bold value represents the statistic, and the corresponding p-value is italicized. A positive sign for the statistic value indicates that the mean value of the corresponding DI (or differential DI) is observed to be higher in magnitude among the control subjects, and a negative sign indicates that they are higher among the subjects with pFC lesions. Statistically insignificant features ($p >= 0.05$) are masked to avoid their contribution to the final analysis. Fig. 5 shows the significant sub-networks for the two populations (control and patients with pFC lesions) that show a greater change in DI from the pretrial to WM encoding phase.



(a) Heatmap of Welch's t-test results for DI features of the WM encoding phase.



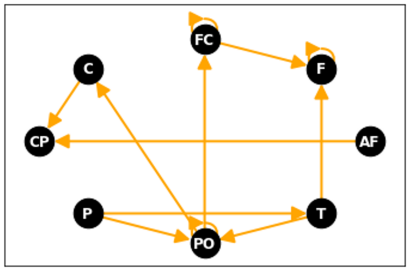
(b) Heatmap of Welch's t-test results for differential DI features of the WM encoding phase from the baseline.

Fig. 4: Statistical comparison of DI and differential DI features for control subjects vs. subjects with pFC lesions.

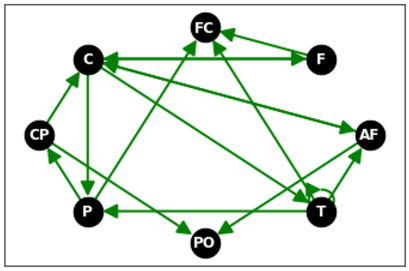
4. DISCUSSION

The main region-wise observations from the statistical test are discussed below:

- The control group exhibits a higher inflow of DI to F, FC, C, and CP regions, while subjects with pFC lesions show a higher DI inflow to the other regions.
- Although there is a significantly higher inflow of information to the AF region for subjects with pFC lesions, information flow from AF is predominantly seen only towards P and PO regions among the subjects with pFC lesions. DI to other regions from AF is relatively higher among the control subjects.
- When compared to the baseline, the differential DI inflow from PO regions is higher among the control, and that from the T region is higher among the subjects with pFC lesions. This result corroborates our previous finding of increased mean betweenness of PO region among controls and T region among subjects with pFC lesions.
- Though the directional information flow for AF→C, F→FC, F→C, F→C, T→FC, P→FC, and P→CP are seen to be higher among the control subjects, these features showed a higher change among subjects with pFC lesions when compared to baseline. This may indicate that these neural pathways are generally active with high information flow among controls but only get activated among subjects with pFC lesions during memory encoding. Similarly, a typically active elevated information flow to PO from P, PO, and T regions exists among subjects with pFC lesions that are only activated during the memory encoding phase among control subjects.
- The neural pathways AF→CP, C→CP, F↔FC, PO→FC, and PO→C show higher information flow exclusively among the control subjects, and C→P, C→T, CP→PO, and T→P exclusively show a higher DI among subjects with pFC.



(a) Sub-graph with differential DI edges that are higher among control subjects.



(b) Sub-graph with differential DI edges that are higher among subjects with pFC lesions.

Fig. 5: Region-wise sub-networks with statistically significant higher differential DI flow among control subjects (5a) an patients with pFC lesions (5b).

5. CONCLUSION

A multi-subject group classification method of subjects with pFC lesions from healthy control using DI inferred effective connectivity is discussed in this paper. We have demonstrated that differential DI are viable features to manage inter-subject variability that typically affects any group-level analysis. Subnetwork group analysis revealed critical differences in directional information flow across the two classes using scalp EEG recorded during WM tasks. Subjects with pFC lesions exhibited a reduced outflow of information from the AF region. Specific neural pathways that are generally active with high information flow for one class only get activated for the other class during memory encoding. Finally, some neural pathways are generally active with high information flow solely for one of the classes. Further work with a much larger data set is needed to generalize the findings and validate our results. A larger data set can also make the use of deep-learning models feasible for its analysis. Future work on identifying directional neural pathways for memory encoding and retrieval among subjects with memory disorders like dementia and Alzheimer's disease may assist in developing prospective treatment procedures.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by [14]. Ethical approval was not required, as confirmed by the license attached with the open-access data.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

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

[2] Robin G Morris, "Working memory in alzheimer-type dementia.," *Neuropsychology*, vol. 8, no. 4, pp. 544, 1994.

[3] 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.

[4] 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.

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

[6] David Friedman and Ray Johnson Jr, "Event-related potential (erp) studies of memory encoding and retrieval: A selective review," *Microscopy research and technique*, vol. 51, no. 1, pp. 6–28, 2000.

[7] 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.

[8] 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.

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

[10] Rakesh Malladi, Giridhar Kalamangalam, Nitin Tandon, and Behnaam Aazhang, "Identifying seizure onset zone from the causal connectivity inferred using directed information," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1267–1283, 2016.

[11] 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.

[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] Sai Sanjay Balaji and Keshab K Parhi, "Classifying subjects with pfc lesions from healthy controls during working memory encoding via graph convolutional networks," in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER), to appear*. IEEE, 2023.

[14] Elizabeth L. Johnson (2017), "64-channel human scalp eeg from 14 unilateral pfc patients and 20 healthy controls performing a lateralized visuospatial working memory task," 2017, CRCNS.org, DOI: http://dx.doi.org/10.6080/K0ZC811B.

[15] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[16] Tingting Xu, Kathryn R Cullen, Bryon Mueller, Mindy W Schreiner, Kelvin O Lim, S Charles Schulz, and Keshab K Parhi, "Network analysis of functional brain connectivity in borderline personality disorder using resting-state fmri," *NeuroImage: Clinical*, vol. 11, pp. 302–315, 2016.

[17] Tingting Xu, Massoud Stephane, and Keshab K Parhi, "Abnormal neural oscillations in schizophrenia assessed by spectral power ratio of meg during word processing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 11, pp. 1148–1158, 2016.

[18] Bhaskar Sen, Kathryn R Cullen, and Keshab K Parhi, "Classification of adolescent major depressive disorder via static and dynamic connectivity," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2604–2614, 2020.

[19] Mohammad Reza Mohammadi, Ali Khaleghi, Ali Moti Nasrabadi, Safa Rafieivand, Moslem Begol, and Hadi Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," *Biomedical Engineering Letters*, vol. 6, no. 2, pp. 66–73, 2016.

[20] Zisheng Zhang and Keshab K Parhi, "Muse: Minimum uncertainty and sample elimination based binary feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 9, pp. 1750–1764, 2018.

[21] Rene J Huster, Sergey M Plis, and Vince D Calhoun, "Group-level component analyses of eeg: validation and evaluation," *Frontiers in neuroscience*, vol. 9, pp. 254, 2015.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "SMOTE: synthetic minority oversampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.