CAPTURING GROUP VARIABILITY USING IVA: A SIMULATION STUDY AND GRAPH-THEORETICAL ANALYSIS

Sai Ma[1], *Ronald Phlypo*[1], *Vince D. Calhoun*[2] *and Tülay Adalı*[1]

[1]University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250
[2]The Mind Research Network and University of New Mexico, Dept. of ECE, Albuquerque, NM 87131

ABSTRACT

When applied to functional magnetic resonance imaging (fMRI) data, independent vector analysis (IVA) provides superior performance in capturing subject variability within one group, as compared to the widely used group independent component analysis (ICA) approach. However, the effectiveness of IVA algorithms in preserving variability between different groups of subjects has not been studied yet, although it is of great interest in most fMRI studies, especially for identifying biomarkers for diagnosis of mental disorders. In this paper, we introduce a methodology that uses graph-theoretical analysis and statistical analysis for assessing the ability of IVA algorithms to capture group variability. We generate multi-subject fMRI-like datasets with increasing spatial variability for a selected component between two groups and compare a robust IVA algorithm to group ICA approach. Our experimental results show that IVA can successfully preserve group variability, indicating its potential in extracting biomarkers across groups of subjects in fMRI analysis.

Index Terms— IVA, ICA, multi-subject fMRI-like data, group variability, graph-theoretical analysis

1. INTRODUCTION

Data-driven methods, such as independent component analysis (ICA), have proven very useful for analysis of functional magnetic resonance imaging (fMRI) data. To achieve a group analysis by ICA for multi-subject fMRI data, a widely used procedure [1, 2] is to reshape 4D data and treat the spatial dimensions as a single dimension first, then concatenate the images from individual subjects in time, followed by principal component analysis (PCA) applied on a group level to reduce data dimension; after a single ICA decomposition, the independent components for individual subjects can be obtained by a back-reconstruction stage [3]. However, since the group-level PCA step attempts to find a common signal space for all subjects, this group ICA method using temporal concatenation may impose restrictions on signal distribution, suggesting difficulties in capturing the inter-subject variability [4, 5] and possibly for group variability as well.

A more effective solution for group fMRI analysis is to use independent vector analysis (IVA), which can concurrently extract independent components from multiple datasets (e.g., subjects) [6, 7]. In IVA, the components from a dataset are assumed to be maximally independent of each other, as in ICA; more importantly, dependence between the given components from all datasets is maximized simultaneously in IVA, which cannot be achieved by separate ICA of each dataset. Without performing the group-level PCA, IVA for fMRI demonstrates the capability to preserve signals that show spatial variability in their hemodynamic responses across a group of subjects [8].

However, the effectiveness of IVA algorithms in capturing spatial variability between different groups of subjects has not been studied to date, which is of great interest in various group fMRI studies, e.g., comparison between healthy controls and patients with mental disorder to identify potential biomarkers for diagnosis and treatment.

In this work, based on graph-theoretical analysis and statistical measures we introduce a methodology for assessing the ability of IVA algorithms to capture group variability. We generate multi-subject fMRI-like datasets using a recently developed simulation toolbox [9]. To simulate a typical difference between the healthy and patient groups found in many fMRI studies, i.e., the brain network volume changes [10, 11], we introduce increasing spread variability in one selected component between two groups of subjects during the data generation stage. We then compare the performance of an effective and robust IVA algorithm with the group ICA approach in preserving the introduced group variability.

2. BACKGROUND

2.1. Independent vector analysis

To introduce IVA for fMRI analysis, consider multiple datasets for K subjects, each denoted as $\mathbf{x}^{[k]}, k = 1 \dots K$, generated from the following linear mixture model

$$\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}, k = 1, \dots, K$$

*Research supported in part by the NSF grant NSF-IIS 1017718.

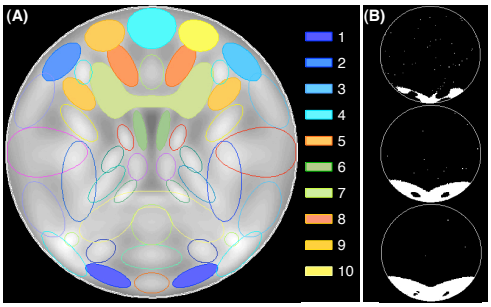

Fig. 1. (A) Selection of ten original components, labeled by different colors. We select this set of components to reduce spatial overlap between component #1 and other components. (B) Mask (corrected by FDR=0.05) of two-sample t-map for original component #1, to which we introduce group variability, in Cases 3–5; then we use this mask to generate ROC curve for two-sample t-map of the estimated component.

where superscript $[\cdot]$ is the subject index. For the kth subject, the component vector is denoted as $\mathbf{s}^{[k]} = \left[s_1^{[k]}, \ldots, s_n^{[k]}, \ldots, s_N^{[k]}\right]^T$, where the subscript is the component index and N is the total number of components. The nth source component vector (SCV) is constructed by taking each of the nth component from all K subjects, i.e., $\mathbf{s}_n = \left[s_n^{[1]}, \ldots, s_n^{[k]}, \ldots, s_n^{[K]}\right]^T$.

The goal of IVA is then to find K demixing matrices $\mathbf{W}^{[k]}, k = 1, \ldots, K$ and the corresponding component vector estimates for each subject, $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, such that the estimated SCVs, given by $\mathbf{y}_n = \left[y_n^{[1]}, \ldots, y_n^{[K]}\right]^T, n = 1, \ldots, N$, are maximally independent of each other. For the nth SCV $\mathbf{y}_n$, the element $y_n^{[k]} = \left(\mathbf{w}_n^{[k]}\right)^T \mathbf{x}^{[k]}$ and $\left(\mathbf{w}_n^{[k]}\right)^T$ is the nth row of $\mathbf{W}^{[k]}$. IVA decomposition can be achieved by minimizing the mutual information among SCVs, i.e., minimizing cost function $\mathcal{I}_{\text{IVA}} = \sum_{n=1}^{N} \mathcal{H}(\mathbf{y}_n) - \mathcal{H}(\mathbf{y}_1, \ldots, \mathbf{y}_N) = \sum_{n=1}^{N} \left(\sum_{k=1}^{K} \mathcal{H}(y_n^{[k]}) - \mathcal{I}(\mathbf{y}_n)\right) - \sum_{k=1}^{K} \log\left|\det \mathbf{W}^{[k]}\right| - C$, where $\mathcal{H}$ is entropy, $\mathcal{I}(\mathbf{y}_n) = \mathcal{I}\left(y_n^{[1]}; \ldots; y_n^{[K]}\right) = E\left\{\log \frac{p(\mathbf{y}_n)}{p(y_n^{[1]}) \ldots p(y_n^{[K]})}\right\}$ is the mutual information (MI) within the nth SCV and C is a constant term that depends only on $\mathbf{x}^{[k]}, k = 1, \ldots, K$. This cost function can be minimized using a parametric model for the SCV distribution $p(\mathbf{y}_n)$ (e.g., multivariate Gaussian or Kotz-type distribution [12]), and an iterative optimization technique (e.g., gradient descent method). Therefore, by minimizing this cost function the dependence among components within each SCV, $\mathcal{I}(\mathbf{y}_n)$, is maximized concurrently.

We use an effective IVA implementation, namely IVA-GL, achieved by initializing IVA-L [6] with a solution from IVA-G [7]. IVA-L is an IVA algorithm with multivariate Laplace prior for SCVs [6] and IVA-G algorithm assumes multivariate Gaussian for SCV distribution instead [7]. Both IVA-L and IVA-G follow the above mixing model and cost function. In IVA-L, the components within each SCV are assumed to be second-order uncorrelated and hence only higher-order statistics are exploited; while IVA-G exploits only second-order statistical information by multivariate Gaussian assumption. Combining second-order and higher-order statistics, IVA-GL can yield more robust joint blind source separation than using IVA-L or IVA-G alone [7].

Table 1: Simulation parameters for all subjects in five cases, where $\mathcal{U}$ and $\mathcal{N}$ denote uniform and Gaussian distributions. Comp., component set; Trans., translation; Rot., rotation.

Group	Case	Comp.	Trans.	Rot.	Spread
Group 1	Case 1	#1	$\mathcal{N}(0,2)$	$\mathcal{N}(0,3)$	$\mathcal{U}(.3,1)$
	Case 2				$\mathcal{U}(.3,1.8)$
	Case 3				$\mathcal{U}(.3,3.3)$
	Case 4				$\mathcal{U}(.3,5.3)$
	Case 5				$\mathcal{U}(.3,8.3)$
	Case 1–5	#2–10	$\mathcal{N}(0,2)$	$\mathcal{N}(0,3)$	$\mathcal{U}(.3,1)$
Group 2	Case 1–5	#1–10	$\mathcal{N}(0,2)$	$\mathcal{N}(0,3)$	$\mathcal{U}(.3,1)$

2.2. Graph-theoretical analysis

An effective way to evaluate the performance of IVA and group ICA approaches in capturing group variability is to use graph-theoretical analysis, as applied in many fMRI studies on groups of subjects [13].

To calculate graph-theoretical metrics, we construct an undirected graph for each of the K subjects. In the kth graph, a node represents an estimated component from the kth subject, $y_n^{[k]}, n = 1, \ldots, N$; an edge connecting two nodes, $y_n^{[k]}$ and $y_m^{[k]}, n \neq m$, represents the normalized MI between them [14, 15], defined as $\sqrt{1 - \exp\left(-2\mathcal{I}(y_n^{[k]}; y_m^{[k]})\right)} \in [0, 1)$ [16], where $\mathcal{I}(\cdot;\cdot)$ is the MI estimated by kernel density method [17]. A threshold, defined as the percentage of the strongest edges retained, gradually increases to generate a number of graphs with different connection densities. At each threshold, the average clustering coefficient to quantify the local cliquishness of nodes in the graph is calculated as $\frac{1}{N}\sum_{n=1}^{N} E_n/(N_n(N_n-1)/2)$ [18], where N_n is the number of edges connected to the nth node and E_n is the number of pairs of the nth node's neighbors that are connected to each other. The obtained value of this metric is also normalized by the corresponding measure from the comparable random graph with the same number of nodes and edges as the observed graph. This metric takes into account relationship of the component that we introduce group variability with all the other components and in our experiments, we expect group differences in this metric when group variability is high.

3. SIMULATION AND STATISTICAL ANALYSIS

3.1. FMRI-like data generation

We generate multi-subject fMRI-like data using a new simulation toolbox, SimTB [9]. Following the linear mixing model, this toolbox controls the generation of 2D spatial components

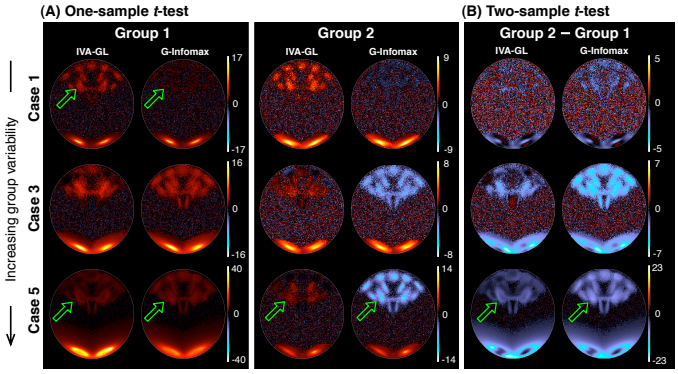

Fig. 2. Statistical testing of component #1 in three cases: (A) one-sample t-test for two groups and (B) two-sample t-test between two groups.

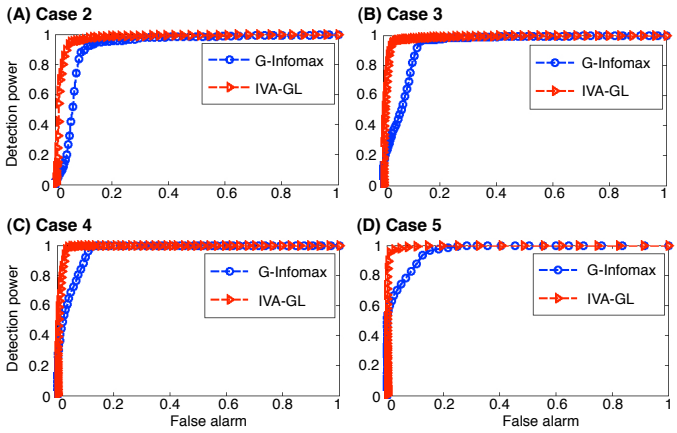

Fig. 3. ROC curve for two-sample t-map of the estimated component #1 for IVA-GL and G-Infomax in Cases 2–5.

and time courses by a selected number of parameters. For example, the slices of individual components across all subjects within one group can be rotated, translated, contracted or expanded based on the distributions of relevant parameters.

Total of 10 original components are used, including visual, frontal, pre-central, ventricle, and anterior white matter, as shown in Fig. 1. (A). Each spatial component contains 148×148 voxels. Gaussian noise with a variance corresponding to a standard contrast-to-noise ratio (CNR) of 2 is added to each component. Time courses are simulated as the convolution of "neural" events with a canonical hemodynamic response function and scaled to have a peak-to-peak range of one. We specify the event magnitude of each time course and control the correlation between all possible pairs of original time courses to be around a small absolute value (0.2) to ensure successful identification of each underlying component.

Various fMRI studies have found significant group differences in the brain network volumes between the healthy and diseased brains. For example, cerebral gray matter volumes are smaller; while ventricle and cerebrospinal fluid volumes are larger in schizophrenia patients than in healthy controls [10, 11]. To simulate this typical group difference in the selected brain network, we generate five cases with expected increasing spatial spread for component #1 in Group 1 (12 subjects) and fix the spread distribution for this component in Group 2 (12 subjects) for all cases. Hence, the functional volume difference in component #1 between two groups increases across these cases. For components #2–10, we use the same set of parameters in both groups and we expect no significant group differences in these components. The simulation parameters for five cases are shown in Table 1.

3.2. Statistical analysis

To evaluate how well IVA and group ICA preserve the group variability in the selected component #1, several statistical measures are calculated. Between two simulated groups, we perform two-sample t-test on the estimated component #1 to obtain t-map, where the higher the voxel value the more significant the group difference in this voxel is. For the original component #1, we also obtain the two-sample t-map and then threshold it (corrected by a false discovery rate (FDR) of 0.05) to create a mask, representing the ground truth of significant differences between two groups, as shown in Fig. 1. (B). Using this mask, we generate the receiver operator characteristic (ROC) curve for the two-sample t-map of the estimated component #1. The number of voxels within the mask (N_{in}) and the number of voxels outside the mask (N_{out}) are calculated. At each t-value threshold in the t-map of the estimated component, we compute the number of voxels that fall in and outside of the mask (N_{true} and N_{false}). The probability of detection and false alarm are defined as the ratios, $N_{\text{true}}/N_{\text{in}}$ and $N_{\text{false}}/N_{\text{out}}$, respectively. Then the set of these ratios obtained by varying the t-value threshold defines the ROC curve. Additionally, we calculate the normalized MI between the two-sample t-maps derived from the estimated and original components (both are thresholded for FDR = 0.05) to measure the statistical similarity between them.

4. EXPERIMENTAL RESULTS

We evaluate the performance of IVA-GL algorithm in capturing group variability and compare with group ICA algorithm implemented in the group ICA for fMRI toolbox (GIFT, `http://icatb.sourceforge.net`). To match IVA-GL, we use the Infomax algorithm in GIFT with a nonlinearity that corresponds to the Laplace distribution, denoted as G-Infomax in the remainder of the text.

First, we calculate the normalized MI between the estimated components and their corresponding true components. For both IVA-GL and G-Infomax across five cases, more than 95% of estimates from all subjects yield high normalized MI values (> 0.5), hence indicating a good estimation performance. For component #1 to which we introduce group variability, the normalized MI for more than 85% of subjects have

Table 2: The normalized MI between two-sample t-map of the estimated and original component #1. Both t-maps are thresholded for FDR = 0.05. In Case 1, no voxel survives after thresholding. The significance of difference between two algorithms is tested by two-sample t-test of 20 trials of random subject selection ($p < 0.05$ for all four cases).

	Case 1	Case 2	Case 3	Case 4	Case 5
	— increasing group variability→				
IVA-GL	–	0.43	0.56	0.81	0.88
G-Infomax	–	0.37	0.51	0.75	0.83

high value (> 0.6) for both algorithms.

We introduce increasing group variability to component #1 across five cases and after source separation, we perform statistical analysis on this component. For the estimation of component #1, the t-maps from one-sample t-test for two groups are shown in Fig. 2. (A) and the t-maps of two-sample t-test for IVA-GL and G-Infomax are shown in Fig. 2. (B). As expected, the number and t-value of significant voxels located at the bottom (the original region subject to increasing spread variability in component #1) increase as the mean and variance of spread distribution in Group 1 increase. When there is no group variability (Case 1), interference from other components in the estimated component #1 is less for G-Infomax than IVA-GL, potentially resulting from the group-level PCA step in G-Infomax that attempts to find a common signal space for all subjects. However, as group variability in spread increases, the estimation of component #1 from IVA-GL shows less interference from the other components than that estimated by G-Infomax, as shown by arrows in Fig. 2. (B).

The ROC curves for the two-sample t-map of component #1 are shown in Fig. 3. We find that across all cases, IVA-GL shows a better performance than G-Infomax in terms of ROC curve. Also, the ROC performance of IVA-GL improves as group variability increases.

Additionally, we calculate the normalized MI between the thresholded two-sample t-maps of the estimated and original component #1, as shown in Table 2, an indication of how well the group differences in the true regions (represented by the two-sample t-map of the original component) are captured in the estimated component. Compared to G-Infomax, the two-sample t-map derived from IVA-GL is more similar to the ground truth.

The graph-theoretical metric of ten estimated components is shown in Fig. 4. When group variability is high (e.g., Cases 4 and 5), the two simulated groups show larger differences using IVA-GL estimations and thus have higher accuracy for the classification of two groups than using G-Infomax estimations, further indicating the performance of IVA-GL in capturing the group differences.

To verify the results we obtained, we run IVA and group ICA algorithms ten times with random initializations (in IVA-GL, we randomly initialize IVA-G algorithm). For all our

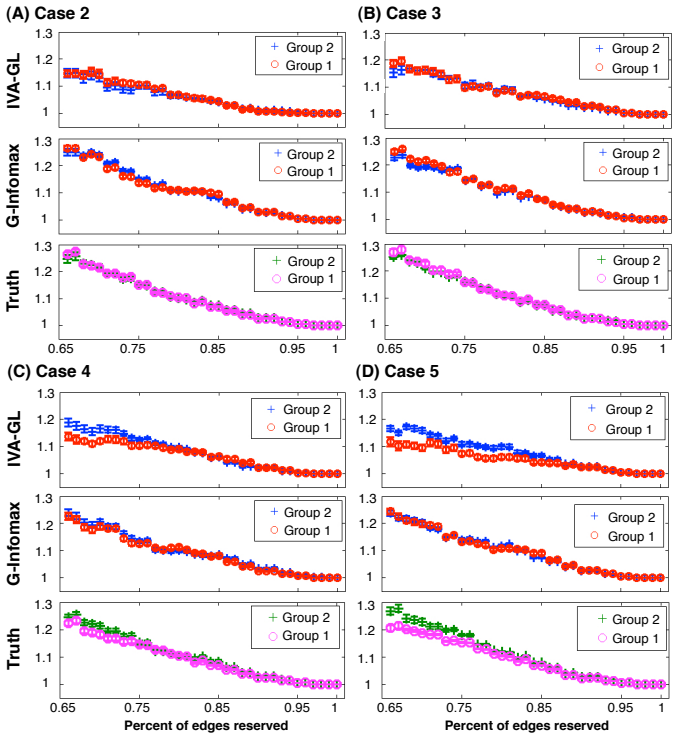

Fig. 4. The average clustering coefficient obtained using ten true components and ten estimated components from IVA-GL and G-Infomax. The introduced group variability in component #1 increases from Case 2 to Case 5.

simulations, we obtain similar results with respect to the performance of the two algorithms in capturing group variability. We also consider an extreme case that we only introduce spread differences between two groups, but within group we do not include subject variability. In such a case, IVA-GL performs better than G-Infomax in terms of graph-theoretical and statistical measures.

5. CONTRIBUTIONS AND CONCLUSIONS

In this paper, we show that IVA not only preserves inter-subject variability with one group of subjects, as shown in [8], but that it also is effective in capturing variability between different groups of subjects, which is especially important for group studies. We generated multi-subject fMRI-like datasets with increasing spatial variability between two groups. Using graph-theoretical analysis and several statistical measures, we compared a robust IVA algorithm to the widely used group ICA approach. Our experimental results show that IVA can successfully preserve group difference when there is group variability and the performance of IVA improves as group variability increases, indicating the promising use of IVA for studying group differences to identify biomarkers for mental disorders and between different conditions, such as different cognitive tasks, age, gender variables or during treatment.

6. REFERENCES

[1] V. D. Calhoun, T. Adalı, G. D. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, pp. 140–151, 2001.

[2] V. D. Calhoun and T. Adalı, "Multi-subject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 60–73, 2011.

[3] E. B. Erhardt, S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adalı, and V. D. Calhoun, "Comparison of multi-subject ICA methods for analysis of fMRI data," *Human Brain Mapping*, vol. 32, no. 12, pp. 2075–2095, 2011.

[4] F. Esposito, T. Scarabino, A. Hyvärinen, J. Himberg, E. Formisano, S. Comani, G. Tedeschi, R. Goebel, E. Seifritz, and F. Di Salle, "Independent component analysis of fMRI group studies by self-organizing clustering," *Neuroimage*, vol. 25, no. 1, pp. 193–205, 2005.

[5] E. A. Allen, E. B. Erhardt, Y. Wei, T. Eichele, and V. D. Calhoun, "Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study," *Neuroimage*, vol. 59, no. 4, pp. 4141–4159, 2011.

[6] J. H. Lee, T. W. Lee, F. A. Jolesz, and S. S. Yoo, "Independent vector analysis (IVA): multivariate approach for fMRI group study," *Neuroimage*, vol. 40, no. 1, pp. 86–109, 2008.

[7] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.

[8] J. T. Dea, M. Anderson, E. A. Allen, V. D. Calhoun, and T. Adalı, "IVA for multi-subject fMRI analysis: A comparative study using a new simulation toolbox," in *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing*, 2011, pp. 1–6.

[9] E. B. Erhardt, E. A. Allen, Y. Wei, T. Eichele, and V. D. Calhoun, "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability," *Neuroimage*, vol. 59, no. 4, pp. 4160–4167, 2011.

[10] H. E. Hulshoff Pol, H. G. Schnack, M. G. Bertens, N. E. van Haren, I. van der Tweel, W. G. Staal, W. F. Baaré, and R. S. Kahn, "Volume changes in gray matter in patients with schizophrenia," *American Journal of Psychiatry*, vol. 159, no. 2, pp. 244–250, 2002.

[11] M. Suzuki, S. Nohara, H. Hagino, K. Kurokawa, T. Yotsutsuji, Y. Kawasaki, T. Takahashi, M. Matsui *et al.*, "Regional changes in brain gray and white matter in patients with schizophrenia demonstrated with voxel-based analysis of MRI," *Schizophrenia Research*, vol. 55, pp. 41–54, 2002.

[12] M. Anderson, G. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the Kotz distribution, and performance bounds," in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'13)*, 2013, accepted.

[13] E. T. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature*, vol. 10, pp. 186–198, 2009.

[14] S. Ma, T. Eichele, N. M. Correa, V. D. Calhoun, and T. Adalı, "Hierarchical and graphical analysis of fMRI network connectivity in healthy and schizophrenic groups," in *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI'11)*, 2011, pp. 1031–1034.

[15] S. Ma, V. D. Calhoun, T. Eichele, W. Du, and T. Adalı, "Modulations of functional connectivity in the healthy and schizophrenia groups during task and rest," *NeuroImage*, vol. 62, no. 3, pp. 1694–1704, 2012.

[16] A. Dionisio, R. Menezes, and D. A. Mendes, "Mutual information: a measure of dependency for nonlinear time series," *Physica A: Statistical Mechanics and its Applications*, vol. 344, no. 1, pp. 326–329, 2004.

[17] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[18] Y. Liu, M. Liang, Y. Zhou, Y. He, Y. Hao, M. Song, C. Yu, H. Liu *et al.*, "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, no. 4, pp. 945–961, 2008.