Citation:
H. Yang, M. A. B. S. Akhonda, F. Ghayem, Q. Long, V. D. Calhoun and T. Adali, "Independent Vector Analysis Based Subgroup Identification from Multisubject fMRI Data," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1471-1475, doi: 10.1109/ICASSP43922.2022.9747224.

DOI:
https://doi.org/10.1109/ICASSP43922.2022.9747224

INDEPENDENT VECTOR ANALYSIS BASED SUBGROUP IDENTIFICATION FROM MULTISUBJECT FMRI DATA

H. Yang, M. A. B. S. Akhonda*, F. Ghayem*, Q. Long*, V. D. Calhoun**, T. Adali**

* Dept. of CSEE, University of Maryland Baltimore County, Baltimore, USA
** Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, USA

ABSTRACT

Identification of homogeneous subgroups of subjects plays a key role in the study of precision medicine. While there are a number of approaches based on the clustering of low-level features such as behavioral variables, work that makes use of fully multivariate nature of medical imaging data is very limited. Given that the individual variability in brain functional networks obtained from functional magnetic resonance imaging (fMRI) data is noted as being both significant and consistent like fingerprints, its use provides a particularly appealing approach to this challenging problem. We present a completely data-driven approach, subgroup identification using independent vector analysis (SI-IVA), which leverages the desirable properties of IVA to uncover the relationship across subjects along with the discovery of subgroup structures revealed by Gershgorin disc theorem. We show that SI-IVA outperforms an eigenanalysis-based approach by simulations. We then apply the method to real fMRI data obtained from patients of during resting state to identify group differences in multiple relevant brain regions including primary somatosensory and motor cortex, which demonstrates that SI-IVA provides interpretable and meaningful results.

***Index Terms*—** Subgroup identification, IVA, Gershgorin disc, Eigenanalysis, Precision medicine

1. INTRODUCTION

Identification of homogeneous subgroups of patients serves as a core step in precision medicine, which aims to tailor medical treatments to the individual patient who has been classified into subpopulations. Based on different subgroups, treatments can be concentrated on those who will benefit, sparing side effects for those who will not [1]. In the clinical field, heterogeneity of patients is a challenging problem for mental disorders such as autism, bipolar disorder, and schizophrenia [2–5]. There are a number of approaches that perform clustering of behavioral variables, clinical, cognitive or other related scores to define subgroups [6, 7], however, work that takes into account the fully multivariate nature of medical imaging data is very limited. Recent work [8] shows that the individual variability in brain functional networks from fMRI data is both significant and consistent on identifying subjects. This makes use of fMRI data on subgroup identification problem very appealing.

A subgroup study using fMRI data can be done via clustering of subjects based on the correlation structures of their functional network activities. A two-step procedure is used in [9] to partition subjects by first extracting subjects' functional connectivity (FC) patterns followed by clustering of those FCs. One limitation of this method is that the final subgroups are found based on clustering of FC patterns obtained through separate independent component analyses (ICA) for each subject. Separate analyses do not fully take the multivariate information across the subject datasets into account. Another recent study, independent vector analysis (IVA) for common subspace extraction (IVA-CS) [10], identifies subgroups by jointly analyzing multi-subject data using IVA. By taking into account the dependence across multiple datasets, IVA generalizes ICA to jointly analyze multiple datasets. Compared with other multi-subject ICA algorithms like Group ICA [11], IVA can more effectively capture inter-subject variability [12]. Using multivariate Gaussian density model, IVA can effectively reveal the correlation structure of the latent variables in multisubject data, which has been leveraged in IVA-CS. Although IVA-CS is a promising approach for the subgroup identification problem, it relies heavily on user-defined parameters. One way to alleviate this issue is to apply eigenanalysis to covariance matrices as proposed in [13, 14]. These papers focus on complete model identification, *i.e.,* identifying the number of all correlated signals across multiple datasets, and as a result, their computational complexity grows significantly with the number of datasets, making them impractical when there are more than a couple of datasets. Obviously, for the subgroup identification problem what we are interested in is a method that can effectively work with a large number of datasets, which correspond to each subject's fMRI data. In this paper, we leverage the strengths of these two approaches, use of IVA with the multivariate Gaussian model (IVA-G) to reliably reveal the correlation structure of latent variables across large number of datasets [10], and use of eigenanalysis of covariance matrices to identify the highly correlated subgroups revealed by IVA-G as in [13, 14].

Hence, we present a new approach, subgroup identification using independent vector analysis (SI-IVA), to automatically identify subgroups from multi-subject fMRI data without the need for user-defined threshold selection for the covariance matrices. By defining a source component vector (SCV), IVA takes the dependence (correlation in the case of IVA-G) across the datasets into account. In addition, IVA-G enables reliable estimation of the correlation structure through strong identifiability guarantees, which is not the case with the multiset extension of canonical correlation analysis, MCCA [15–17]. SI-IVA leverages these advantages of IVA and uses it as a guide for subgroup identification. Since the correlation structure of the underlying datasets are captured through SCVs, SI-IVA is fundamentally different from the aforementioned two-step procedure in [9]. Inspired by [13, 14], we propose to use eigenanalysis with Geshgorin disc theorem (GDT) on the SCV covariance matrices for

This work was supported in part by NSF grants CCF 1618551, NCS 1631838 and HRD 2112455, and NIH grants R01MH123610 and R01MH118695. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF).

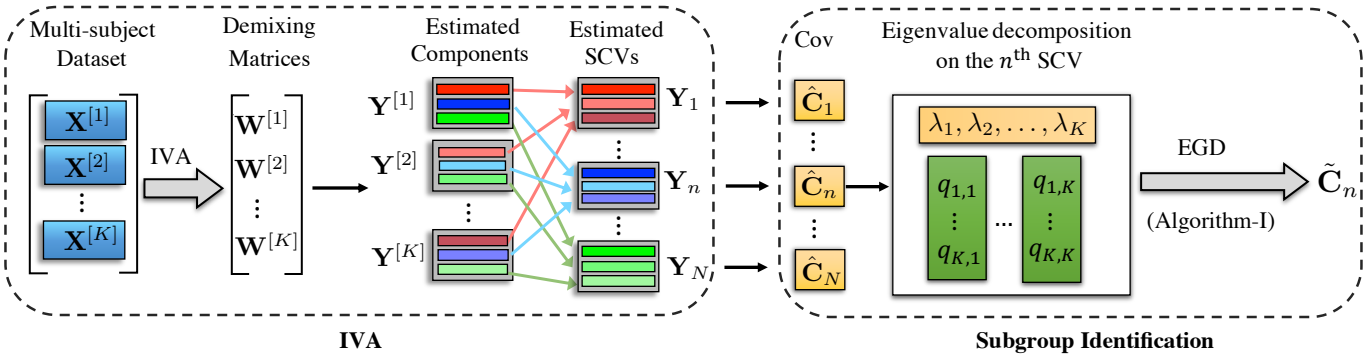


Fig. 1: Block diagram of the proposed method for subgroup identification, SI-IVA. The first step is to apply IVA-G on a multi-subject dataset. Then, eigenanalysis based on Gershgorin disc (EGD) is applied to each SCV covariance matrix. Based on the eigenvalues that are greater than the smallest radius of Gershgorin discs and the corresponding eigenvectors, the number of subgroups and the subjects that belong to the subgroups are identified.

subgroup identification. GDT ensures that the eigenvalues of a matrix are included in sets such as a given disc [18]. We show that the number of eigenvalues that are located outside the smallest disc provides an effective way for identifying the number of subgroups. Since GDT provides a range of eigenvalues for the covariance matrix of each SCV, predefined thresholds are no longer needed. Comparing with the most recent work that is also based on the use of IVA, IVA-CS [10], the proposed method is (i) fully data-driven ; (ii) identifies subgroups based on each individual SCV, which allows us to determine the contribution of each component that define subgroups, also improving interpretability. We test the performance of the proposed method on both simulated and real data. Using simulation results, we show that SI-IVA outperforms the eigenanalysis-based approach as in [13, 14] with respect to the correctly estimated number of subgroups. With real fMRI data, we demonstrate that the subgroups identified by SI-IVA are interpretable and meaningful.

The rest of this paper is organized as follows: the background for IVA and eigenanalysis-based approach are presented in Section 2. The details of the proposed method are in Section 3. Simulation results are presented in Section 4, followed by the conclusion in Section 5.

2. BACKGROUND

This section details the two parts that compose SI-IVA, independent vector analysis and eigenanalsis-based approach for subgroup identification.

2.1. Independent Vector Analysis (IVA)

Given K datasets, IVA models each dataset as a mixture of N independent sources. At a sample index v, the IVA generative model the kth dataset is given as:

$$\mathbf{x}^{[k]}(v) = \mathbf{A}^{[k]}\mathbf{s}^{[k]}(v), 1 \leq k \leq K \tag{1}$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}$ are invertible mixing matrices. In addition to the assumption of independence among sources within a dataset, IVA models the dependence across datasets by defining an SCV as $\mathbf{s}_n(v) = [s_n^{[1]}(v), \ldots, s_n^{[K]}(v)]^\top \in \mathbb{R}^K, 1 \leq n \leq N$, consisting of the nth source component $s_n^{[k]}(v)$ from each of the K datasets. The goal of IVA is to estimate K demixing matrices that minimize the mutual information among the SCVs, which can be achieved by the following cost function

$$\mathcal{I}_{\text{IVA}}(\boldsymbol{\mathcal{W}}) = \sum_{n=1}^{N} \mathcal{H}(\mathbf{y}_n) - \sum_{k=1}^{K} \log|\det(\mathbf{W}^{[k]})| \tag{2}$$

where $\boldsymbol{\mathcal{W}} = \{\mathbf{W}^{[1]}, \mathbf{W}^{[2]}, \ldots, \mathbf{W}^{[K]}\}$ are the K datasets' demixing matrices, $\mathbf{y}_n$ is the estimated SCV, and $\mathcal{H}(\mathbf{y}_n) = -\mathbb{E}\{\log p_n(\mathbf{y}_n)\}$ is the (differential) entropy of $\mathbf{y}_n$ and $p_n(\cdot)$ is the multivariate probability density function (pdf) of the nth SCV. Minimization of (2) is equivalent to maximization of likelihood [16]. The estimated sources of each dataset are given by $\mathbf{y}^{[k]}(v) = \mathbf{W}^{[k]}\mathbf{x}^{[k]}(v)$ for $k = 1, \ldots, K$. With V samples of the observed data, the source estimates are given as the matrix equation $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$, with $\mathbf{Y}^{[k]}, \mathbf{X}^{[k]} \in \mathbb{R}^{N\times V}$, and the estimate source vector is given as $\mathbf{Y}_n = [\mathbf{y}_n^{[1]}, \mathbf{y}_n^{[2]}, \ldots, \mathbf{y}_n^{[K]}]^\top \in \mathbb{R}^{K\times V}$. Here, $\mathbf{y}_n^{[k]}$ is the estimated component from the kth dataset belonging to the nth SCV. The IVA generative model is shown in Fig. 1.

Depending on the choice of pdf of SCVs, IVA can take second-order statistics (SOS) and/or higher-order statistics (HOS) into account. IVA with multivariate Gaussian distribution (IVA-G) assumes that the sources in an SCV are multivariate Gaussian distributed. IVA-G only takes SOS into account, and can be shown to have a positive definite Hessian matrix of the objective function [17]. Recent work on subgroup identification, IVA-CS [10], shows that because of the strong identifiability condition of IVA-G, *i.e*, the ability to uniquely identify source signals under very general conditions, the dependence structure among datasets are well preserved with IVA-G. IVA-CS is also a pioneer work of using IVA to identify subgroups. However, the need of setting multiple user-defined parameters limits its reproducibility of the method. Besides , IVA-CS identifies subgroups based on the average of multiple SCV covariance matrices, which makes it hard to evaluate the contributions of different SCVs. As we discuss next, by analyzing each individual SCV covariance matrix separately, the contribution of different components can be easily revealed, and the user-defined parameters required in IVA-CS can be automatically determined by looking at specific eigenvalues and eigenvectors of the SCV covariance matrices.

2.2. Subgroup identification by eigenanalysis

We assume that, all components are zero mean and unit variance so that the definition of covariance and correlation coincide. In order to

apply the eigenanalysis-based approach on the covariance matrices to subgroup identification, the following two assumptions are made:

- Components are uncorrelated within a dataset and correlated only among the corresponding components across datasets. This is automatically satisfied in the latent space when IVA is applied.
- All correlations are transitive so that SCV covariance matrices have block structures.

We define the element on the k_1^{th} row, k_2^{th} column of the nth SCV covariance matrix as $\rho_n^{[k_1,k_2]} = \mathbb{E}\{s_n^{[k1]} s_n^{[k2]}\}$, *i.e.,* the correlation coefficient between k_1 and k_2 for the nth SCV. Based on eigenvalue decomposition, we factor the sample covariance matrix into $\hat{\mathbf{C}}_n = (1/V)\mathbf{Y}_n\mathbf{Y}_n^\top = \mathbf{QDQ}^\top$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements λ_n^i are the ith eigenvalue of $\hat{\mathbf{C}}_n$, and $\mathbf{q}_n^i \in \mathbb{R}^K$ is the corresponding eigenvector, the ith column of $\mathbf{Q} \in \mathbb{R}^{K\times K}$.

Subgroup identification problem aims at identifying a group of subjects across which the components in the nth estimated SCV are correlated. The correlation of components within an SCV means the subjects have similar activated functional networks. Since the components in each SCV come from individual subjects, the index of the correlated components can be used to identify subjects that belong to the same subgroup. Based on [14], for each SCV covariance matrix $\hat{\mathbf{C}}_n$, the number of eigenvalues that are greater than one and the corresponding eigenvectors can be used to identify the number of subgroups and the index of subjects that belong to each subgroup. In the rest of this paper, we refer this method as eigenanalysis based on hard thresholding (EHT).

However, without thresholding the covariance matrix, the number of eigenvalues that are greater than one does not accurately reflect the number of subgroups. One example is showed in Fig. 2b, where the middle plot is the eigen-profile of the SCV covariance matrix that is on the left side. The simulation set up is discussed in Section 4. We can see that 4 eigenvalues are greater than 1, which implies 4 subgroups while there are only two. For practical fMRI data, which typically has a low signal-to-noise ratio, the eigen-profile of $\hat{\mathbf{C}}_n$ typically has a smooth transition causing eigenanalysis-based approaches to overdetermine the number of subgroups. A suboptimal solution is to set a user-defined value τ, however its choice is difficult and might vary for each $\hat{\mathbf{C}}_n$. As we discuss in Section 3, this problem can be addressed by making use of GDT as we propose.

3. PROPOSED METHOD

Given the fact that the subgroup structures are well preserved in IVA-G analysis, the first step of our new method, SI-IVA, is to apply IVA-G on a multi-subject dataset. Let $\hat{\mathbf{C}}_n$ be the sample covariance matrix of the nth SCV. Based on this SCV, we define subgroups as the diagonal blocks that form $\hat{\mathbf{C}}_n$, where each block points out the correlations between the subjects in the corresponding subgroup. Now, consider after certain permutations, the covariance matrix $\hat{\mathbf{C}}_n$ has $m > 1$ block-diagonal structures as $\hat{\mathbf{C}}_n = \text{blkdiag}(\mathbf{G}_n^1, \ldots, \mathbf{G}_n^m, \mathbf{I})$, where $\mathbf{G}_n \in \mathbb{R}^{g\times g}$ represents the correlation coefficients between the g subjects within the subgroup associated with the nth SCV. Then, in order to identify the number of subgroups, m, eigenanalysis based on Gershgorin disc (EGD) is applied on each $\hat{\mathbf{C}}_n$. To this end, we first define Gershgorin discs on the sample covariance matrix $\hat{\mathbf{C}}_n$.

Let $R_i = \sum_{j\neq i} |\rho_n^{[i,j]}|$ be the sum of the absolute values of the non-diagonal entries in the ith row of $\hat{\mathbf{C}}_n$. A Gershgorin disc is a closed disc centered at $\rho_n^{[i,i]}$ with radius R_i, $\{z \in \mathbb{R} : |z - \rho_n^{[i,i]}| \leq R_i\}$. Examples of Gershgorin disc can be found in Fig. 2. For the normalized covariance matrix with unit variance, the diagonal entries are 1, $\rho_n^{[i,i]} = 1$. Since $\hat{\mathbf{C}}_n$ is a symmetric matrix, the eigenvalues are always real. So the center of the discs of $\hat{\mathbf{C}}_n$ are always located at (1,0). The radius R_i of each disc represents the energy of the ith row of $\hat{\mathbf{C}}_n$. The eigenvalues of $\hat{\mathbf{C}}_n$ locate in the union of Gershgorin discs, $\text{eig}(\hat{\mathbf{C}}_n) \in \bigcup_{i=1}^K \{z \in \mathbb{R} : |z - \rho_n^{[i,i]}| \leq R_i\}$. Now, consider R_{min} to be the smallest radius which reflects the lowest energy among the rows of the covariance matrix. To identify the number of subgroups, we find the number of the eigenvalues of $\hat{\mathbf{C}}_n$ that are located outside the smallest disc, *i.e.* greater than $R_{\text{min}} + 1$. These eigenvalues can be interpreted as the energy of the blocks (subgroups) in $\hat{\mathbf{C}}_n$, which contain the largest percentages of the whole energy in $\hat{\mathbf{C}}_n$.

The last step of SI-IVA is to identify the indexes of the associated subjects that belong to each subgroup. To this aim, we apply K-means clustering [19] on the eigenvectors corresponding to the eigenvalues that are outside the smallest disc. According to the identified subgroup indexes, a permutation matrix $\mathbf{P}_n$ can be formed to sort the columns of $\hat{\mathbf{C}}_n$. The hidden subgroup structure is revealed by $\tilde{\mathbf{C}}_n = \mathbf{P}_n\hat{\mathbf{C}}_n$. The process of implementing SI-IVA is summarized in Algorithm-1 and the flowchart of SI-IVA is shown in Fig. 1. We note that in Algorithm-1 the eigenvalues are sorted as $\lambda_1 > \lambda_2 > \cdots > \lambda_N$.

Algorithm 1 eigenanalysis based on Gershgorin disc (EGD)

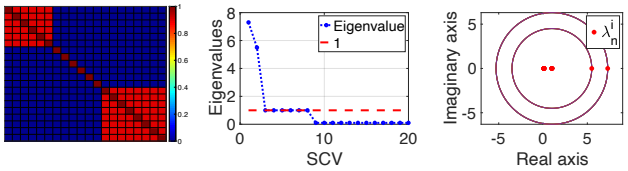
Inputs: $\hat{\mathbf{C}}_n$: estimated covariance matrix of the nth SCV
$M = 0$
$\lambda_n^i, \mathbf{q}_n^i \leftarrow$ sort(eig($\hat{\mathbf{C}}_n$))
$R_i \leftarrow \sum_{j\neq i} |\rho_n^{[i,j]}|$
if $\lambda_n^i > R_{\text{min}} + 1$ **then**
    the number of subgroup: $M \leftarrow M + 1$
    corresponding eigenvector: $\mathbf{q}_n^i$
end if
for $m = 1, 2, \cdots, M$ **do**
    $\mathbf{P}_n \leftarrow$ K-means($\mathbf{q}_n^m$, 2)
end for
$\tilde{\mathbf{C}}_n \leftarrow \mathbf{P}_n\hat{\mathbf{C}}_n$
Output: $\tilde{\mathbf{C}}_n$

4. EXPERIMENTAL RESULTS

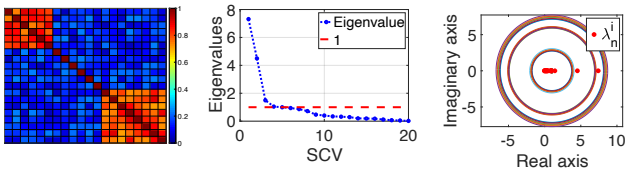
In this section, we first compare our proposed method SI-IVA with the eigenanalysis approach based on hard thresholding (EHT) [14] on synthetic data. Then we test SI-IVA with fMRI data.

4.1. Application to synthetic data

To compare the prediction accuracy regarding to the number of subgroups from EGD and eigenanalysis-based approach, we run these two algorithms 100 times on a set of simulated data. All SCVs used for simulation are generated from multivariate generalized Gaussian distribution (MGGD) [20], which has super-Gaussian marginals for $\beta < 1$, sub-Gaussian for $\beta > 1$ and is Gaussian when $\beta = 1$. The shape parameter β is randomly selected from the interval [0.1, 0.8], which is a good match for fMRI data. The correlation values for correlated and uncorrelated sources are $\rho_c = \mathcal{U} \sim [0.7, 0.9]$ and $\rho_d = \mathcal{U} \sim [0.05, 0.25]$ separately. A total of $N = 20$ SCVs are generated with $V = 10000$ voxels and $K = 15$ datasets. The first 5 SCVs are simulated as components that are highly correlated across all subjects with correlation value as ρ_c. The second group of 5 SCV



(a) $\rho_c = 0.9$, $\rho_d = 0$ separately.



(b) $\rho_c = \mathcal{U} \sim [0.7, 0.9]$, $\rho_d = \mathcal{U} \sim [0.05, 0.25]$.

Fig. 2: Application of EHT and EGD to simulated data. The first column shows the generated SCV covariance matrices, $\hat{\mathbf{C}}_n$. The second column is the eigen-profile of $\hat{\mathbf{C}}_n$. The blues dots are the eigenvalues λ_n^i. We look for eigenvalues that are greater than 1 (above the red dash line) to represent the number of block structures (subgroups) in $\hat{\mathbf{C}}_n$. The third column are the results of applying eigenanalysis based on Gershgorin disc (EGD) to $\hat{\mathbf{C}}_n$. The red dots and the circles are the actual eigenvalues λ_n^i and the Gershgorin discs of $\hat{\mathbf{C}}_n$. The correlation values for the correlated and uncorrelated subjects are (a): $\rho_c = 0.9$, $\rho_d = 0$, (b): $\rho_c = \mathcal{U} \sim [0.7, 0.9]$, $\rho_d = \mathcal{U} \sim [0.05, 0.25]$. When ρ_c, ρ_d are specified values, the number of subgroups detected by eigenanalysis and EGD are the same as EHT. When the values of ρ_c, ρ_d are within a range, EHT shows 4 eigenvalues are greater than 1, *i.e.,* 4 subgroup are detected, but EGD shows 2 eigenvalues are located outside the smallest circle, *i.e.,* 2 subgroup are detected.

covariance matrices have block structures with high/low correlation values as ρ_c and ρ_d respectively. One example can be found in Fig. 2b. The block structure represents subgroups, *i.e.*, some subjects are highly correlated, but others are not. The third group of 5 SCV covariance matrices are simulated for the components that are varied from subject to subject with low correlation values ρ_d. As it is shown in Fig. 2 (a), when the off-diagonal elements of $\hat{\mathbf{C}}_n$ are exactly zero, the number of subgroups identified by EGD is the same as EHT. As it is seen Fig. 2b, our method is still able to detect subgroup number correctly when the off-diagonal elements' values of $\hat{\mathbf{C}}_n$ are noisy, where EHT approach fails. We run this simulation 100 times. In all, our method was able to identify the correct subgroup number, while the EHT failed in detecting the right subgroup number.

4.2. Application to multi-subject resting-state fMRI data

We use resting-state fMRI data collected from 88 schizophrenia patients (SZs) and 91 healthy controls (HCs). Fifty of the patients were randomly selected for implementing the proposed method. The data is from the Center of Biomedical Research Excellence (COBRE) [21–23], and is available to download from (`https://coins.trendscenter.org/`). Participants were instructed to keep their eyes open during the scan and stare passively at a central fixation cross. Each subject data consists of 144 volumes and $53 \times 63 \times 46$ voxels.

IVA-G is applied to 50 patients that are randomly selected and 85 SCVs are estimated as in [10]. The value at each voxel of the estimated source is Z-scored before any calculation of metrics, so the covariance and correlation coincide. To achieve better reproducibility of the results, we run IVA-G 10 times and select the most

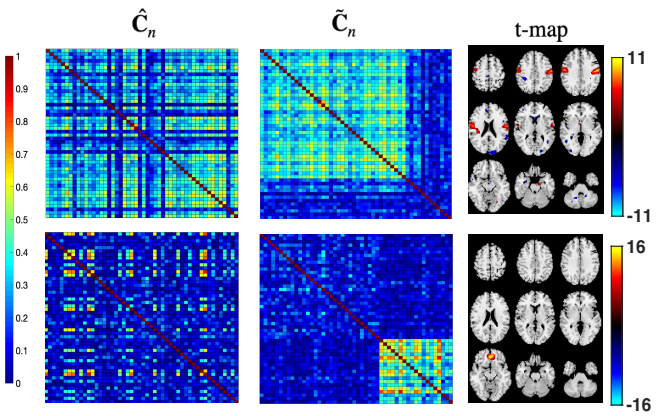


Fig. 3: Results from SI-IVA. The underlying subgroup structure in $\hat{\mathbf{C}}_n$ is revealed in $\tilde{\mathbf{C}}_n$. T-maps include the brain areas that show significant group difference. The first row shows the subgroup identified from SCV 50, where primary somatosensory and motor cortex show significant differences between subgroups ($p = 1.6 \times 10^{-4}$). The second row is from SCV 62, which is a particular case that a very small area of activation shows significant differences ($p = 2 \times 10^{-4}$).

consistent run using cross intersymbol interference (Cross-ISI) measure [24]. Applying the EGD to each SCV covariance matrix returns the number of eigenvalues that are located outside the smallest disc, *i.e.,* the subgroup number. K-means clustering is performed on the corresponding eigenvectors to identify the subject indexes that belong to different subgroups. Based on the subject indexes, the subgroup structures are reveal by permuting the original covariance matrix.

To compare the spatial activation patterns of resting-state networks (RSNs) across subgroups in each SCV, a two-sample t-test is performed on the activation value at each voxel of the spatial map across the subjects within each subgroup. The areas that show significant group differences ($p \leq 0.05$) are highlighted in the t-maps. False discovery rate (FDR) correction is conducted throughout all comparisons. Fig. 3 shows the SCV covariance matrices before and after SI-IVA, Brain regions that show significant subgroup differences include primary somatosensory and motor cortex, and a small area of activation that shows significant differences between subgroups.

5. CONCLUSION AND DISCUSSION

A new method of subgroup identification using IVA, SI-IVA, is proposed to automatically divide subjects into subgroups based on the correlation of their functional network activities. Comparing with the most recent work IVA-CS, which is also based on the use of IVA, SI-IVA is fully data-driven and identifies subgroups based on each SCV, which allows us to determine the contribution of each SCV that defines subgroups, which is important for interpreting the results. The simulation results show that EGD is more robust in identifying subgroup numbers on low signal-to-noise ratio data. It should be added that, the level of correlation that constitutes "commonality" for subgroup definition affects the performance, as well as the accuracy of IVA estimates which might over/under estimate correlation values when there is not enough sample support. In addition, the properties of GDT needs to be established to provide theoretical guarantees for its performance.

6. REFERENCES

[1] N. R. Council et al., Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease," 2011.

[2] Y. Shao, M. Cuccaro, E. Hauser, K. Raiford, M. Menold, C. Wolpert, S. Ravan, L. Elston, K. Decena, S. Donnelly, et al., Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes," *The American Journal of Human Genetics*, vol. 72, no. 3, pp. 539–548, 2003.

[3] M. Dekker, V. Bonifati, and C. Van Duijn, Parkinson's disease: piecing together a genetic jigsaw," *Brain*, vol. 126, no. 8, pp. 1722–1733, 2003.

[4] W. K. Scott, E. R. Hauser, D. E. Schmechel, K. A. Welsh-Bohmer, G. W. Small, A. D. Roses, A. M. Saunders, J. R. Gilbert, J. M. Vance, J. L. Haines, et al., Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22," *The American Journal of Human Genetics*, vol. 73, no. 5, pp. 1041–1051, 2003.

[5] D. B. Dwyer, C. Cabral, L. Kambeitz-Ilankovic, R. Sanfelici, J. Kambeitz, V. Calhoun, P. Falkai, C. Pantelis, E. Meisenzahl, and N. Koutsouleris, Brain subtyping enhances the neuroanatomical discrimination of schizophrenia," *Schizophrenia bulletin*, vol. 44, no. 5, pp. 1060–1069, 2018.

[6] O. Veatch, J. Veenstra-VanderWeele, M. Potter, M. Pericak-Vance, and J. Haines, Genetically meaningful phenotypic subgroups in autism spectrum disorders," *Genes, Brain and Behavior*, vol. 13, no. 3, pp. 276–285, 2014.

[7] V. Bitsika, C. Sharpley, and S. Orapeleng, An exploratory analysis of the use of cognitive, adaptive and behavioural indices for cluster analysis of ASD subgroups," *Journal of Intellectual Disability Research*, vol. 52, no. 11, pp. 973–985, 2008.

[8] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable, Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity," *Nature neuroscience*, vol. 18, no. 11, pp. 1664–1671, 2015.

[9] J. Durieux and T. F. Wilderjans, Partitioning subjects based on high-dimensional fMRI data: comparison of several clustering methods and studying the influence of ICA data reduction in big data," *Behaviormetrika*, vol. 46, no. 2, pp. 271–311, 2019.

[10] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adali, Independent vector analysis for common subspace analysis: Application to multi-subject fMRI data yields meaningful subgroups of schizophrenia," *NeuroImage*, vol. 216, pp. 116872, 2020.

[11] V. D. Calhoun, T. Adali, 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.

[12] A. M. Michael, M. Anderson, R. L. Miller, T. Adalı, and V. D. Calhoun, Preserving subject variability in group fMRI analysis: performance evaluation of GICA vs. IVA," *Frontiers in systems neuroscience*, vol. 8, pp. 106, 2014.

[13] T. Hasija, T. Marrinan, C. Lameiro, and P. J. Schreier, Determining the dimension and structure of the subspace correlated across multiple data sets," *Signal Processing*, vol. 176, pp. 107613, 2020.

[14] M. Akhonda, B. Gabrielson, V. D. Calhoun, and T. Adali, Complete Model Identification Using Independent Vector Analysis: Application to the Fusion of Task FMRI Data," in *2021 IEEE Data Science and Learning Workshop (DSLW)*. IEEE, 2021, pp. 1–6.

[15] J. R. Kettenring, Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[16] T. Adali, M. Anderson, and G.-S. Fu, Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[17] M. Anderson, T. Adali, 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, 2011.

[18] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge university press, 2012.

[19] J. MacQueen et al., Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, 1967, vol. 1, pp. 281–297.

[20] E. Gómez, M. Gomez-Viilegas, and J. M. Marín, A multivariate generalization of the power exponential family of distributions," *Communications in Statistics-Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[21] A. Scott, W. Courtney, D. Wood, R. De la Garza, S. Lane, R. Wang, M. King, J. Roberts, J. A. Turner, and V. D. Calhoun, COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets," *Frontiers in neuroinformatics*, vol. 5, pp. 33, 2011.

[22] M. S. Cetin, F. Christensen, C. C. Abbott, J. M. Stephen, A. R. Mayer, J. M. Cañive, J. R. Bustillo, G. D. Pearlson, and V. D. Calhoun, Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia," *Neuroimage*, vol. 97, pp. 117–126, 2014.

[23] C. Aine, H. J. Bockholt, J. R. Bustillo, J. M. Cañive, A. Caprihan, C. Gasparovic, F. M. Hanlon, J. M. Houck, R. E. Jung, J. Lauriello, et al., Multimodal neuroimaging in schizophrenia: description and dissemination," *Neuroinformatics*, vol. 15, no. 4, pp. 343–364, 2017.

[24] Q. Long, C. Jia, Z. Boukouvalas, B. Gabrielson, D. Emge, and T. Adali, Consistent run selection for independent component analysis: Application to fMRI analysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2581–2585.