

CONSTRAINED INDEPENDENT COMPONENT ANALYSIS BASED ON ENTROPY BOUND MINIMIZATION FOR SUBGROUP IDENTIFICATION FROM MULTISUBJECT FMRI DATA

H. Yang, F. Ghayem*, B. Gabrielson*, M. A. B. S. Akhonda*, 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 subgroups of subjects homogeneous functional networks is a key step for precision medicine. Independent vector analysis (IVA) is shown to be effective for this task, however, it has a substantial computing cost. We propose a constrained independent component analysis algorithm based on minimizing the entropy bound (c-EBM) to overcome the computational complexity limitation of IVA. A set of spatial maps used as constraints provides a connection across the datasets, provides alignment across subject-wise ICA analyses and serves as a foundation for subgroup identification. The approach makes use of the available prior knowledge while allowing flexible density modeling without an orthogonality requirement for the demixing matrix. Synthetic data and large scale multi-subject resting state fMRI data have both been used to evaluate the performance of the new algorithm, c-EBM. The findings demonstrate that c-EBM is adaptable in terms of various settings for the constraint parameter on the synthetic data. With multi-subject resting state fMRI data, c-EBM can effectively identify subgroups and discover meaningful brain networks that show significant group differences between subgroups.

Index Terms— subgroup identification, constrained ICA, multi-subject data, resting state fMRI, precision medicine

1. INTRODUCTION

Precision medicine aims to provide individualized medical care for individual patients by precisely identifying subgroups based on key characteristics of patients. Numerous subtypes from various disease domains have been proposed, including those for neuropsychiatric conditions like autism [1], bipolar disorder [2], schizophrenia [3], and major depression [4]. Patients can be divided into heterogeneous subpopulations (subgroups) based on the corresponding biomarkers of these subtypes. Classifying neuropsychiatric patients is a significant difficulty in precision medicine because the etiology of neuropsychiatric illnesses is uncertain [5–8]. Focusing on clustering subjective data, such as behavioral characteristics, clinical, cognitive, or other related scores, is one approach to solving this challenge [9, 10]. However, the relationship between subtypes and post hoc descriptions is not well understood, hence the success of such approaches is limited. Functional magnetic resonance imaging (fMRI) has been widely used to study functional connectivity of the brain, and these networks are identified to be unique as fingerprints [11, 12].

This work was supported in part by NSF-NCS 1631838, NSF 2112455, and NIH grants R01 MH118695, R01 MH123610, R01 AG073949. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF): hpcf.umbc.edu.

Identification of putative biomarkers of neuropsychiatric disorders using independent component analysis (ICA) is shown to be effective [13–19]. Group ICA (GICA) [20] is a commonly used method for multi-subject data analysis using ICA by vertically concatenating the subject datasets and performing one ICA on the group data. However, GICA assumes common spatial maps across subject and only incorporates subject variability through the back reconstruction step. On the other hand, extention of ICA to multiple datasets, independent vector analysis (IVA) is shown to better preserve subject variability [19, 21–24], and has been shown to successfully identify subgroups of patients [25]. However, IVA is computationally complex as it requires computation of multivariate density whose dimensionality depends on the number of datasets (subjects), and hence, e.g., in [25] only a subset of 50 patients are used in the analysis.

Constrained ICA (c-ICA) can be used to analyze multi-subject data by making use of constraints to link multiple subjects. The connection across subjects is established by the aligned components through the constraint and the overall source separation performance can improve with the provided prior information, such as in the form of spatial maps that model functional connectivity networks [26–30]. As c-ICA is applied to individual subjects, it avoids the penalty in the cost of IVA caused by over-parameterization while maintaining subject variability compared with GICA [31]. A current approach to constrained ICA makes use of orthogonality constraint on the demixing matrix and uses fixed nonlinearity limiting its performance [28–30]. The orthogonality constraint limits the solution space, also limiting the performance [32].

To address the aforementioned problems, we introduce constrained EBM (c-EBM), which is based on minimizing the bound on source entropy [33]. In conjunction with allowing subject datasets fully interact via constraints, our approach also provides a flexible match to source probability distributions and is able to decouple the demixing matrix without imposing an orthogonality constraint. Simulated datasets and large scale multi-subject resting state fMRI data are utilized to evaluate the effectiveness of the c-EBM. On simulated data, the comparison with the recent constrained ICA method, multi-objective optimization ICA with a reference (MOO-ICAR) [30] reveals that c-EBM is more adaptable to various constraint parameter settings. With multi-subject resting state fMRI data from 88 subjects, we show that c-EBM provides promising performance in the identification of subgroups of subjects and discovering meaningful brain networks including visual, primary sensory, and motor area that show significant group differences between subgroups. Compared with MOO-ICAR, the results from c-EBM have lower average normalized mutual information between estimated sources, indicating better separation performance. In addition, the work demonstrates the replicability of subgroup results in [25] for a

larger number of patients.

The rest of this paper is organized as follows: the background for ICA is presented in Section 2. The details of the proposed method c-EBM and its application on subgroup identification are in Section 3 and Section 4. Simulation results are presented in Section 5, followed by the discussion in Section 6.

2. BACKGROUND

Let $\mathbf{x}(v) = [x_1(v), \dots, x_N(v)]^\top$ be the observation vector (e.g. from a single subject) modeled by $\mathbf{x}(v) = \mathbf{A}\mathbf{s}(v)$ at sample index v , where $\mathbf{s}(v) = [s_1(v), \dots, s_N(v)]^\top$ are N statistically independent zero mean latent sources and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is an invertible mixing matrix. The estimation of the dataset's latent sources $\mathbf{y}(v) = [y_1(v), \dots, y_N(v)]^\top$ can be achieved by $\mathbf{y}(v) = \mathbf{W}\mathbf{x}(v)$, where \mathbf{W} is the demixing matrix. For simplicity, the sample index v is suppressed in the rest of this paper. We can write the cost function $\mathcal{J}(\mathbf{W})$ as mutual information among source estimates y_n for $n = 1, \dots, N$, which is a function of the demixing matrix \mathbf{W} as follows

$$\mathcal{J}(\mathbf{W}) = \sum_{n=1}^N \mathcal{H}(y_n) - \log|\det(\mathbf{W})| - \mathcal{H}(\mathbf{x}), \quad (1)$$

where $\mathcal{H}(y_n)$ represents the (differential) entropy of the n^{th} estimated source y_n , $\mathcal{H}(\mathbf{x})$ is the entropy of \mathbf{x} and hence is constant with respect to \mathbf{W} , and $\log|\det(\mathbf{W})|$ acts as a regularization term. Now in order to identify the demixing matrix \mathbf{W} for the estimation of the latent sources, we minimize $\mathcal{J}(\mathbf{W})$ with respect to \mathbf{W} . To minimize $\mathcal{J}(\mathbf{W})$ with respect to \mathbf{W} , it is required to know the distribution of the latent sources y_n in order to calculate $\mathcal{H}(y_n)$. In the ICA algorithm based on entropy bound minimization (EBM) [33], entropy is estimated by estimating an upper bound among several measuring functions that come from various distributions. The flexible density estimation strategy of EBM provides a more accurate and reliable source separation performance as compared with many other ICA algorithms [33, 34]. For this reason, in this paper we choose ICA-EBM for our constrained ICA approach.

An orthogonal \mathbf{W} helps decouple the estimation of density for each source, hence providing better model match. However, the orthogonality constraint of \mathbf{W} limits the solution space of \mathbf{W} . An effective decoupling method [35] without constraining \mathbf{W} to be orthogonal is applied to (1) to divide the original problem into a series of sub-problems that minimize the mutual information among the estimated sources with respect to each row vector \mathbf{w}_n , corresponding to each individual source y_n . After decoupling, (1) can be written as

$$\mathcal{J}_n(\mathbf{w}_n) = \mathcal{H}(y_n) - \log|\mathbf{d}_n^\top \mathbf{w}_n| + C, \quad (2)$$

where C is a constant that is independent of \mathbf{w}_n and \mathbf{d}_n^\top is a unit vector that is perpendicular to all the rows of \mathbf{W} except \mathbf{w}_n [35]. Based on the entropy estimator proposed in [33], the cost function of ICA-EBM can be written as

$$\mathcal{J}_n(\mathbf{w}_n) = -O_{m(n)}\{E[G_{m(n)}(y_n)]\} - \log|\mathbf{d}_n^\top \mathbf{w}_n| + C_1, \quad (3)$$

where $O_{m(n)}\{E[G_{m(n)}(y_n)]\}$ is the negentropy of the n^{th} estimated source, C_1 is a constant that is independent of \mathbf{w}_n , and $G_{m(n)}$ is a measure function selected from M measure functions for the n^{th} source. We note that $m = 1, \dots, M$ is a function of n since different measuring functions can be used for different sources. In this paper, we consider $M = 4$ measuring functions $G(x)$: $x^4, \frac{|x|}{1+|x|},$

$\frac{x|x|}{10+|x|}, \frac{x}{1+x^2}$. Also, we note that the negentropy function $O(\cdot)$ is calculated beforehand and saved as a table.

3. CONSTRAINED EBM

With the decoupled cost function (3), constraints can be applied for individual sources without assuming orthogonality of \mathbf{W} . For a given constraint function h_n , (3) is optimized subject to

$$h_n(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) = \theta_n - \epsilon(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) \leq 0, \quad (4)$$

where $\epsilon(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n)$ measures the distance between estimated source and its corresponding reference, and θ_n is a constraint parameter that controls the tolerance of $\mathbf{w}_n^\top \mathbf{x}$ to deviate from \mathbf{r}_n . To measure distance we use Pearson correlation $\epsilon(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) = \text{corr}((\mathbf{w}_n^\top \mathbf{x})^\top, \mathbf{r}_n)$. With the definition of Pearson correlation, $\epsilon(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n)$ and θ_n are restricted between 0 and 1. A higher value of θ_n imposes a stronger constraint to the estimated source, which increases the similarity between the estimated sources and the references, and thus decreases the variability across subjects. On the other hand, a lower value of θ_n allows subject variability to be preserved but may not fully take advantage of the prior information.

To incorporate the inequality constraint (4) into the cost function (3), we use the augmented Lagrangian approach. We define a slack variable z as $h_n(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) + z^2 = 0$, and replace the inequality constraint with an equality. With an optimization process as in [29], the cost function of c-EBM is written as

$$\begin{aligned} \mathcal{J}_n^c(\mathbf{w}_n, \mu_n) &= \mathcal{J}_n(\mathbf{w}_n) \\ &- \frac{1}{2\gamma_n} \left[\max\{0, [\gamma_n h_n(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) + \mu_n]\}^2 - \mu_n^2 \right], \end{aligned} \quad (5)$$

where μ_n is a Lagrangian multiplier and $\gamma_n \in \mathbb{R}_*^+$ is a learning parameter. Gradient descent approach is used for $\min_{\mathbf{w}_n} \mathcal{J}_n^c(\mathbf{w}_n)$, where updates are written as

$$\begin{aligned} \Delta \mathbf{w}_n &\propto \frac{\partial \mathcal{J}_n^c(\mathbf{w}_n)}{\partial \mathbf{w}_n} \\ &= -O'_{m(n)}\{E[G_{m(n)}(y_n)]\}E[g_{m(n)}(y_n)\mathbf{x}] \\ &\quad - \frac{\mathbf{d}_n}{\mathbf{d}_n^\top \mathbf{w}_n} - h'_n(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n)\mu_n \mathbf{r}_n^\top, \end{aligned} \quad (6)$$

where in $\Delta \mathbf{w}_n = \mathbf{w}_n^{i+1} - \mathbf{w}_n^i$, i denotes the iteration index, and $O(\cdot)', g_{m(n)}$ and h'_n are the first order derivatives of $O(\cdot)$, $G_{m(n)}(\cdot)$ and $h_n(\cdot)$ with respect to \mathbf{w}_n , respectively. In each iteration, the Lagrange multiplier μ_n is updated by

$$\mu_n \leftarrow \max\{0, [\gamma_n h_n(\mathbf{w}_n^\top \mathbf{x}, \mathbf{r}_n) + \mu_n]\}. \quad (7)$$

4. SUBGROUP IDENTIFICATION

A homogeneous subgroup can be defined as a group of subjects that have higher correlation among themselves across multiple functional networks than the rest of the subjects. With the constraint included, c-EBM is able to align the components across subject datasets with respect to the corresponding reference \mathbf{r}_n , in our case, the activation map of a functional network. For a given set of observation $\mathbf{X} \in \mathbb{R}^{N \times V}$, the estimates are $\mathbf{Y} = \mathbf{WX}$. To evaluate the similarity of a given functional network among K subjects, we define source component vector (SCV) as $\mathbf{s}_n = [s_n^{[1]}, \dots, s_n^{[K]}]^\top$, i.e., by concatenating the n^{th} source $s_n^{[k]}$, $1 \leq k \leq K$, from each of the K subject dataset. The n^{th} SCV summarizes the spatial activation map corresponding to the n^{th} source, for example visual network,

for all K subjects. The sample covariance matrix of the n^{th} SCV, $\hat{\mathbf{C}}_n = (1/(V-1))\mathbf{Y}_n \mathbf{Y}_n^\top$, provides the correlation information among K subjects for a given functional network.

To identify homogeneous subgroups, a two-step subgroup identification approach described in [25] is implemented: step I, clustering, is applied to the estimated covariance matrix of SCVs, $\hat{\mathbf{C}}_n$, by k-means clustering. The goal is to group the SCVs into sets of brain networks that have similar activation patterns across all subjects; step II, modularization, maximizes the modularity of the mean covariance matrix of each cluster to reveal the subgroup structure as shown in Fig. 3(a). The subgroup identification results reflect the intragroup similarity within a subgroup and intergroup difference across subgroups in terms of the spatial activation patterns of subjects' functional networks.

5. EXPERIMENTAL RESULTS

In this section, the performance of the proposed method c-EBM is examined first using synthetic data, then with multi-subject resting state fMRI data for identification of subgroups.

5.1. Simulation results using synthetic data

The synthetic data are generated for $K = 5$ datasets (subjects), in which $N = 10$ sources with $V = 10^4$ samples are generated from multivariate generalized Gaussian distribution (MGGD) [36], which is super-Gaussian when the shape parameter $\beta < 1$, sub-Gaussian for $\beta > 1$, and Gaussian when $\beta = 1$. The shape parameter β is randomly generated from a uniform distribution $\beta \sim \mathcal{U}(0.2, 0.8)$ for 6 super-Gaussian sources and $\beta \sim \mathcal{U}(2, 10)$ for 4 sub-Gaussian sources. The correlation among sources within an SCV is uniformly distributed in the range [0.8, 0.95].

The correlation between the references and the real sources is ρ_{true} , which is changed from 0.1 to 0.9 with 0.1 as step-size aiming to test the performance of c-EBM with references that have different quality levels. Reference signals are generated in such a way that they reflect ρ_{true} correlation with the average component of the simulated SCV. Three algorithms—c-EBM, EBM, and MOO-ICAR—are tested for each scenario. Since MOO-ICAR is designed to be initialized from references, we take into account 5 runs with random initializations only for EBM and c-EBM. The entire process is repeated 200 times. Fig. 1 shows a diagram of the simulation design.

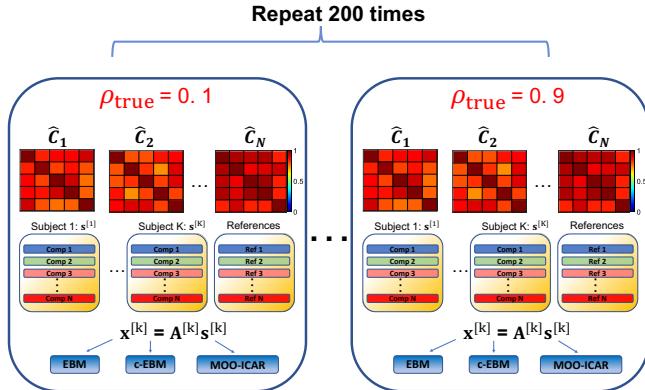


Fig. 1: Diagram summarizing the synthetic data generation

To evaluate the effectiveness of the aforementioned algorithms, we use the correlation between estimated and the real sources, as well as joint inter-symbol-interference (joint-ISI) [37], which is defined as

$$\text{ISI}_{\text{JNT}}(\mathbf{G}^{[1]}, \dots, \mathbf{G}^{[K]}) \triangleq \text{ISI}\left(\frac{1}{K} \sum_{k=1}^K |\mathbf{G}^{[k]}|\right), \quad (8)$$

where

$$\begin{aligned} \text{ISI}(\mathbf{G}) &= \frac{1}{2N(N-1)} \left(\sum_{n=1}^N \left(\frac{\sum_{m=1}^N |g_{nm}|}{\max_p |g_{np}|} - 1 \right) \right. \\ &\quad \left. + \left(\sum_{m=1}^N \left(\frac{\sum_{n=1}^N |g_{nm}|}{\max_p |g_{mp}|} - 1 \right) \right) \right), \end{aligned} \quad (9)$$

with $\mathbf{G}^{[k]} = \mathbf{W}^{[k]} \mathbf{A}^{[k]}$ elements g_{nm} [38, 39], where $\mathbf{A}^{[k]}$ is the true mixing matrix and $\mathbf{W}^{[k]}$ is the estimated demixing matrix. The range of ISI is $0 \leq \text{ISI} \leq 1$, and when the demixing matrix $\mathbf{W}^{[k]}$ is perfectly estimated, $\mathbf{G}^{[k]}$ is an identity matrix subject to permutation and scaling ambiguities. *i.e.*, ISI is 0.

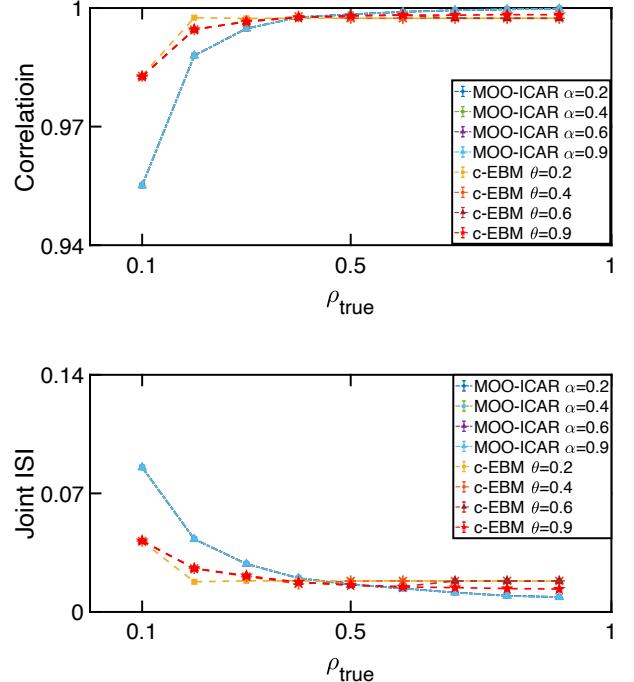


Fig. 2: Top: correlation between the estimated sources and the true sources. Bottom: joint-ISI v.s. ρ_{true} for c-EBM and MOO-ICAR. Both c-EBM and MOO-ICAR separate sources accurately when the quality of reference signals is high, *i.e.*, ρ_{true} is large. When ρ_{true} is low, c-EBM outperforms MOO-ICAR. The constraint tuning parameter for c-EBM and MOO-ICAR are θ and α separately.

We can observe from Fig. 2 that c-EBM performs best in terms of correlation and joint ISI when θ is close to ρ_{true} . Both c-EBM and MOO-ICAR separate sources accurately when the quality of reference signals is high, *i.e.*, ρ_{true} is large. When there are not enough reliable references, c-EBM performs better than MOO-ICAR. The performance of c-EBM can be easily controlled by changing the constraint parameter θ based on the available prior information. However, the performance of MOO-ICAR is not truly affected by the change in constraint parameter α . We observe that the MOO-ICAR results with different tuning parameters overlap and are on top each other. As MOO-ICAR lacks flexibility, its performance is highly dependent on the reference, which might be problematic if the relationship between references and real sources is unclear. Since the components from EBM are not aligned across subjects, EBM has high joint-ISI and low correlation value that is located outside the plot range of Fig. 2 and those values are not affected by ρ_{true} . The average of joint-ISI and correlation over different realizations of EBM are reported as $0.16 \pm 3.5 \times 10^{-6}$ and $0.1 \pm 3.8 \times 10^{-6}$ respectively.

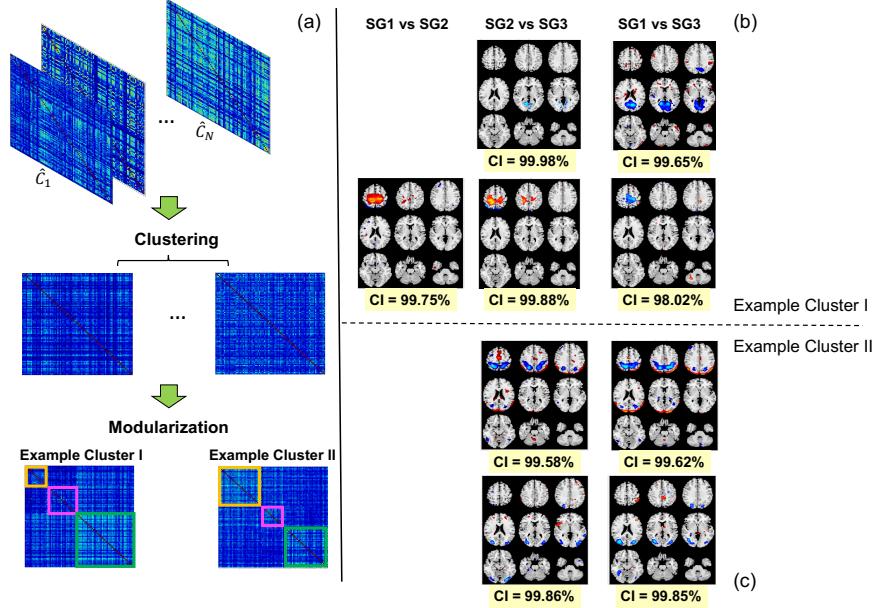


Fig. 3: The subgroup identification process for identifying subgroups of patients. a) Three subgroups are marked by yellow, magenta, and green squares. b) and c) report the resting-state networks (RSNs) showing significant group differences across subgroups from example cluster I and example cluster II, respectively. The confidence intervals (CI) after FDR correction are included in yellow boxes at the bottom of each plot. RSNs that are identified by c-EBM are showing significant group differences in visual, primary sensory and motor areas. These RSNs match with the RSNs reported in [25].

5.2. Experimental results on resting state fMRI data

In this part, the proposed method c-EBM is applied to 88 patients with schizophrenia (SZ) using resting-state fMRI data obtained from the Center of Biomedical Research Excellence (COBRE) [40–42]. The data from each subject consists of 144 volumes and $53 \times 63 \times 46$ voxels. The dataset is available at <https://coins.trendscenter.org/>. We use 53 spatial maps as references collected from NeuroMark pipeline [14].

With 53 as the number of components, which corresponds to the number of references, c-EBM is applied to each subject individually. Once the estimated sources have been retrieved using c-EBM, the value at each voxel of the estimated sources is Z-scored so that the covariance and correlation coincide. By stacking the corresponding components from each subject, we form SCVs and its sample covariance matrix \hat{C}_n . Based on the subgroup identification process introduced in Section 3, followup the modularization stage, three subgroups (SG) were found from each cluster. A two-sample t -test is used to analyze the activation value at each voxel of the spatial map across the subjects within each subgroup to determine whether the spatial activation patterns of resting-state networks (RSNs) show significant group differences across subgroups. False discovery rate (FDR) correction [43] is included in all comparisons. In Fig. 3(b) and (c) the regions that show significant differences ($p\text{-value} \leq 0.05$) between subgroups are highlighted in the t-maps. Also the confidence intervals (CI) after FDR correction are reported. Significant group differences in visual, primary sensory, and motor areas are observed in RSNs identified by c-EBM. These RSNs match with the RSNs reported in [25]. Given the fact that [25] identifies subgroups based on COBRE dataset as well, the overlapping RSNs verifies the ability of c-EBM on identifying subgroups. For MOO-ICAR and c-EBM, the average normalized mutual information [34] between sources across 88 subjects is $0.0155 \pm 1.43 \times 10^{-4}$ and

$0.0066 \pm 2.61 \times 10^{-4}$, respectively. When compared with IVA, c-EBM has a much lower computational complexity. The wall time of implementing IVA is 74 hours and 24 minutes with 50 individuals [44], whereas c-EBM is 5 hours and 4 minutes with 88 subjects. The reported wall time came from running both algorithms on the UMBC High Performance Computing Facility (HPCF).

6. DISCUSSION

Identifying subgroups from large-scale multi-subject fMRI data is a challenging problem and has very recently started to receive attention. In this work, a constrained ICA algorithm based on entropy bound minimization, c-EBM, is proposed to jointly identify subgroups from multi-subject fMRI data. Compared with the most recent subgroup identification approaches based on IVA [25, 44], c-EBM achieves similar performance but with significantly reduced computational time and memory utilization. The simulation findings show that c-EBM performed well in terms of correlation and joint-ICI to separate sources. The flexibility of controlling constraint tuning parameter makes c-EBM robust against the quality of references. Compared with other c-ICA algorithms, *i.e.*, MOO-ICAR [30], c-EBM is able to separate sources accurately when the quality of references is not satisfactory.

The references for c-EBM used in the current study came from the NeuroMark pipeline [14]. Future research can look more closely at how use of different references affect c-EBM results in terms of subgroup identification. The impact of using different ICA algorithms to generate references on the performance of c-EBM also needs to be investigated. Last but not least, while the current c-EBM requires input in the form of a specified constraint parameter, a variant of the method that can adaptively choose a constraint value might be desirable.

7. REFERENCES

- [1] I. Voineagu, X. Wang, P. Johnston, et al., “Transcriptomic analysis of autistic brain reveals convergent molecular pathology,” *Nature*, vol. 474, no. 7351, pp. 380–384, 2011.
- [2] J. Biederman, E. Mick, S. V. Faraone, et al., “Pediatric mania: a developmental subtype of bipolar disorder?,” *Biol. Psychiatry*, vol. 48, no. 6, pp. 458–466, 2000.
- [3] M. T. Tsuang, M. J. Lyons, and S. V. Faraone, “Heterogeneity of schizophrenia,” *BJPsych*, vol. 156, no. 1, pp. 17–26, 1990.
- [4] J. L. Payne, J. T. Palmer, and H. Joffe, “A reproductive subtype of depression: conceptualizing models and moving toward etiology,” *Harv. Rev. Psychiatry*, vol. 17, no. 2, pp. 72–86, 2009.
- [5] Y. Shao, M. Cuccaro, E. Hauser, et al., “Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes,” *Am. J. Hum. Genet.*, vol. 72, no. 3, pp. 539–548, 2003.
- [6] 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.
- [7] W. K. Scott, E. R. Hauser, D. E. Schmeichel, et al., “Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22,” *Am. J. Hum. Genet.*, vol. 73, no. 5, pp. 1041–1051, 2003.
- [8] D. B. Dwyer, C. Cabral, L. Kambeitz-Ilankovic, et al., “Brain subtyping enhances the neuroanatomical discrimination of schizophrenia,” *Schizophr. Bull.*, vol. 44, no. 5, pp. 1060–1069, 2018.
- [9] O. Veatch, J. Veenstra-VanderWeele, M. Potter, et al., “Genetically meaningful phenotypic subgroups in autism spectrum disorders,” *G2B*, vol. 13, no. 3, pp. 276–285, 2014.
- [10] V. Bitsika, C. Sharpley, and S. Orapeling, “An exploratory analysis of the use of cognitive, adaptive and behavioural indices for cluster analysis of ASD subgroups,” *J Intellect Disabil Res*, vol. 52, no. 11, pp. 973–985, 2008.
- [11] A. Jablensky, “The diagnostic concept of schizophrenia: its history, evolution, and future prospects,” *DCNS*, vol. 12, no. 3, pp. 271, 2010.
- [12] V. D. Calhoun, T. Eichele, and G. Pearson, “Functional brain networks in schizophrenia: a review,” *Front. Hum. Neurosci.*, vol. 3, pp. 17, 2009.
- [13] S. Bhinge, Q. Long, V. D. Calhoun, et al., “Spatial dynamic functional connectivity analysis identifies distinctive biomarkers in schizophrenia,” *Front. Neurosci.*, vol. 13, pp. 1006, 2019.
- [14] Y. Du, Z. Fu, J. Sui, et al., “Neuromark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders,” *NeuroImage Clin.*, vol. 28, pp. 102375, 2020.
- [15] V. D. Calhoun and T. Adali, “Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery,” *IEEE Rev Biomed Eng*, vol. 5, pp. 60–73, 2012.
- [16] V. D. Calhoun, P. K. Maciejewski, G. D. Pearson, et al., “Temporal lobe and “default” hemodynamic brain modes discriminate between schizophrenia and bipolar disorder,” *Hum. Brain Mapp.*, vol. 29, no. 11, pp. 1265–1275, 2008.
- [17] C. Sorg, V. Riedl, M. Mühlau, et al., “Selective changes of resting-state networks in individuals at risk for Alzheimer’s disease,” *PNAS*, vol. 104, no. 47, pp. 18760–18765, 2007.
- [18] M. D. Greicius, G. Srivastava, A. L. Reiss, et al., “Default-mode network activity distinguishes Alzheimer’s disease from healthy aging: evidence from functional MRI,” *PNAS*, vol. 101, no. 13, pp. 4637–4642, 2004.
- [19] 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.
- [20] V. D. Calhoun, T. Adali, G. D. Pearson, et al., “A method for making group inferences from functional MRI data using independent component analysis,” *Hum. Brain Mapp.*, vol. 14, no. 3, pp. 140–151, 2001.
- [21] J. Laney, K. P. Westlake, S. Ma, et al., “Capturing subject variability in fMRI data: A graph-theoretical analysis of GICA vs. IVA,” *J. Neurosci. Methods*, vol. 247, pp. 32–40, 2015.
- [22] A. M. Michael, M. Anderson, R. L. Miller, et al., “Preserving subject variability in group fMRI analysis: performance evaluation of GICA vs. IVA,” *Front. Syst. Neurosci.*, vol. 8, pp. 106, 2014.
- [23] J. T. Dea, M. Anderson, E. Allen, et al., “IVA for multi-subject fMRI analysis: A comparative study using a new simulation toolbox,” in *MLSP*. IEEE, 2011, pp. 1–6.
- [24] S. Ma, R. Phlypo, V. D. Calhoun, et al., “Capturing group variability using IVA: a simulation study and graph-theoretical analysis,” in *ICASSP*. IEEE, 2013, pp. 3128–3132.
- [25] Q. Long, S. Bhinge, V. D. Calhoun, et al., “Independent vector analysis for common subspace analysis: Application to multi-subject fMRI data yields meaningful subgroups of schizophrenia,” *NeuroImage*, vol. 216, pp. 116872, 2020.
- [26] M. De Vos, L. De Lathauwer, and S. Van Huffel, “Spatially constrained ICA algorithm with an application in EEG processing,” *Signal Processing*, vol. 91, no. 8, pp. 1963–1972, 2011.
- [27] W. Lu and J. C. Rajapakse, “ICA with reference,” *Neurocomputing*, vol. 69, no. 16–18, pp. 2244–2257, 2006.
- [28] Q. H. Lin, J. Liu, Y.-R. Zheng, et al., “Semiblind spatial ICA of fMRI using spatial constraints,” *Hum. Brain Mapp.*, vol. 31, no. 7, pp. 1076–1088, 2010.
- [29] P. A. Rodriguez, M. Anderson, X.-L. Li, et al., “General non-orthogonal constrained ICA,” *IEEE Trans. Signal Process*, vol. 62, no. 11, pp. 2778–2786, 2014.
- [30] Y. Du and Y. Fan, “Group information guided ICA for fMRI data analysis,” *NeuroImage*, vol. 69, pp. 157–197, 2013.
- [31] M. S. Salman, Y. Du, D. Lin, et al., “Group ICA for identifying biomarkers in schizophrenia: ‘adaptive’ networks via spatially constrained ICA show more sensitivity to group differences than spatio-temporal regression,” *NeuroImage Clin.*, vol. 22, pp. 101747, 2019.
- [32] J. F. Cardoso, “On the performance of orthogonal source separation algorithms,” in *Proceedings of EUSIPCO*. Citeseer, 1994, vol. 94, pp. 776–779.
- [33] X. L. Li and T. Adali, “Independent component analysis by entropy bound minimization,” *IEEE Trans. Signal Process*, vol. 58, no. 10, pp. 5151–5164, 2010.
- [34] Q. Long, S. Bhinge, Y. Levin-Schwartz, et al., “The role of diversity in data-driven analysis of multi-subject fMRI data: Comparison of approaches based on independence and sparsity using global performance metrics,” *Hum. Brain Mapp.*, vol. 40, no. 2, pp. 489–504, 2019.
- [35] M. Anderson, X.-L. Li, P. Rodriguez, et al., “An effective decoupling method for matrix optimization and its application to the ICA problem,” in *ICASSP*. IEEE, 2012, pp. 1885–1888.
- [36] E. Gómez, M. Gomez-Villegas, and J. M. Marín, “A multivariate generalization of the power exponential family of distributions,” *Commun. Stat. - Theory Methods*, vol. 27, no. 3, pp. 589–600, 1998.
- [37] M. Anderson, T. Adali, and X.-L. Li, “Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis,” *IEEE Trans. Signal Process*, vol. 60, no. 4, pp. 1672–1683, 2011.
- [38] S. I. Amari, “Estimating functions of independent component analysis for temporally correlated signals,” *Neural Comput.*, vol. 12, no. 9, pp. 2083–2107, 2000.
- [39] Q. Long, C. Jia, Z. Boukouvalas, et al., “Consistent run selection for independent component analysis: Application to fMRI analysis,” in *ICASSP*. IEEE, 2018, pp. 2581–2585.
- [40] A. Scott, W. Courtney, D. Wood, et al., “COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets,” *Front. Neuroinformatics*, vol. 5, pp. 33, 2011.
- [41] M. S. Cetin, F. Christensen, C. C. Abbott, et al., “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.
- [42] C. Aine, H. J. Bockholt, J. R. Bustillo, et al., “Multimodal neuroimaging in schizophrenia: description and dissemination,” *Neuroinformatics*, vol. 15, no. 4, pp. 343–364, 2017.
- [43] Y. Benjamini and D. Yekutieli, “False discovery rate-adjusted multiple confidence intervals for selected parameters,” *JASA*, vol. 100, no. 469, pp. 71–81, 2005.
- [44] H. Yang, M. Akhonda, F. Ghayem, et al., “Independent vector analysis based subgroup identification from multisubject fMRI data,” in *ICASSP*. IEEE, 2022, pp. 1471–1475.