Dictionary Learning-Based fMRI Data Analysis for Capturing Common and Individual Neural Activation Maps

Rui Jin Krishna K. Dontaraju Seung-Jun Kim M. A. B. S. Akhonda Tülay Adali
Dept. of Computer Science & Electrical Engineering
University of Maryland, Baltimore County
Baltimore, MD 21250
{rjin1,krishna5,sjkim,mo32,adali}@umbc.edu

***Abstract*—A novel dictionary learning (DL) method is proposed to estimate sparse neural activations from multi-subject fMRI data sets. By exploiting the label information such as the patient and the normal healthy groups, the activation maps that are commonly shared across the groups as well as those that can explain the group differences are both captured. The proposed method was tested using real fMRI data sets consisting of schizophrenic subjects and healthy controls. The DL approach not only reproduced most of the maps obtained from the conventional independent component analysis (ICA), but also identified more maps that are significantly group-different, including a number of novel ones that were not revealed by ICA. The stability analysis of the DL method and the correlation analysis with separate neuropsychological test scores further strengthen the validity of our analysis.**

I. Introduction

Functional magnetic resonance imaging (fMRI) has become a representative neuroimaging modality, thanks to its contributions to understanding brain functions and pathologies [1]–[4]. FMRI can reveal the areas of high neural activities non-invasively based on neurovascular coupling [5]. The spatial resolution offered by fMRI is much finer than other existing modalities such as electroencephalography (EEG) and magnetoencephalography (MEG), often providing important complementary views in multi-modal studies [6], [7]. The group analysis of fMRI data can capture common functional networks across multiple subjects [8], [9]. The fMRI analysis is also useful for identifying variations present in different subgroups and individuals. In particular, fMRI studies are instrumental for finding potentially distinguishing biomarkers for a variety of brain diseases [1], [2], [7].

The model-driven approaches for fMRI data analysis typically correlate the time-series data with a hypothesized reference signal such as the hemodynamic response function (HRF), as in the general linear model (GLM) implemented in the statistical parametric mapping (SPM) software [10]. However, the results depend on the reliability of the assumptions made. On the contrary, data-driven approaches make minimal assumptions, rendering the analysis more robust to modeling errors, adaptive to various nuisances in the data, and accommodating to individual traits in multiple data sets.

Blind source separation (BSS) approaches such as independent component analysis (ICA) have been the major data-driven analysis tools for fMRI data [11]. The ICA method aims to extract statistically independent sources from their linear mixture observations. When applied to fMRI data analysis, ICA obtains functionally meaningful spatial maps, without imposing any constraints in the temporal domain [8], [9]. ICA can be extended to the case where multiple data sets corresponding to different subjects are to be processed together. Group ICA concatenates the multi-subject data sets across the time dimension, before performing ICA jointly [12]. The independent vector analysis (IVA) method captures the dependence across the data sets by forming so-called source component vectors (SCVs). Then, the independence across the SCVs and the dependence within each SCV are maximized [13], [14]. Some of the spatial components identified this way were shown to be revealing the differences between the patient and healthy control groups with better performance than the widely used group ICA [13], [15].

More recent approaches include the use of deep learning techniques, which exhibit impressive performance in various machine learning and computer vision applications. A restricted Boltzmann machine (RBM) and a deep belief network (DBN) were employed for fMRI and structural MRI (sMRI) data analysis in [16], and the interpretation of the learned deep network features was attempted using their nonlinear embeddings. Convolutional neural networks (CNNs) were adopted for predicting Alzheimer's disease and autism spectrum disorder (ASD) with good prediction performances [17], [18]. In order to interpret the learned biomarkers, the difference in prediction performance was analyzed when a region of interest (ROI) was corrupted in the input data [18], or a direct sensitivity analysis was performed to the learned network [19]. In summary, while deep learning methods extract useful biomarkers, the interpretation remains challenging, especially as the depth of the network increases.

In this work, a dictionary learning (DL) approach is developed for fMRI data analysis [20], [21]. DL postu-

This work was supported in part by NSF grants 1631838 and 1618551, as well as a UMBC seed grant.

This is a preprint. For the final version, please refer to IEEE Journal of Selected Topics in Signal Processing, DOI: 10.1109/JSTSP.2020.2992430.

lates that each data sample can be represented as a linear combination of a small number of atoms in the dictionary, thus capturing a union-of-subspace structure. Rather than employing a predefined dictionary such as the Fourier or the wavelet basis, DL aims at learning the basis from the data, which can even be overcomplete if the number of measurements is small compared to the dimension of the subspaces. The unsupervised learning method boils down to factorizing the data matrix into a dictionary matrix and a sparse code matrix, while minimizing the reconstruction error. The approach is flexible in that one can incorporate various side information to the learning cost in the form of appropriate regularizers. DL has shown state-of-the-art performance in a variety of applications including image denoising and inpainting, as well as object recognition [22]–[27].

The learned dictionary atoms are often interpretable, which is a critical merit in medical image analysis. In [28], a DL formulation was applied to task fMRI data with a minimum description length (MDL) criterion for determining the sparsity level. It has been shown that the obtained time courses were highly correlated with the canonical HRFs and the spatial activation maps were localized to the relevant brain areas. DL was employed for whole-brain task fMRI data for brain functional network identification in [29]. A hierarchical probabilistic model for brain activity patterns was proposed in the framework of DL for analyzing multi-subject resting-state fMRI data to estimate subject-specific time courses and the spatial maps that are shared across the population [30]. The performances of the DL and ICA methods were compared using various metrics in [31], where it was shown that the time course extracted from DL often exhibits the spectral patterns that match well with actual neural activities.

The DL framework can be used for supervised learning as well, where the dictionary is trained to capture distinctive patterns in the data such that the corresponding sparse codes can be used for predicting the labels for the data [32], [33]. This can be done by simply employing a cost function that augments the reconstruction error with the label prediction error, or by directly seeking a discriminative dictionary via minimizing the prediction error of the classifier as a function of the dictionary.

In the context of fMRI data analysis, one is not necessarily interested in improving the prediction accuracy itself, but rather in capturing detailed neural activities by incorporating available labels in the data sets. For instance, using multi-subject data sets that have group labels available, one can obtain the patterns that are common across the population, as well as the patterns that are representative of the distinct traits in different groups. A dictionary shared across all subjects and the set of subject-specific dictionaries were learned in [34], where incoherence among the learned dictionaries was ensured by penalizing high pairwise correlations. The task fMRI data parcellated to ROIs for schizophrenic patients and normal controls were analyzed in a DL framework in [35], where it was postulated that the schizophrenic group's responses contain sparse subspace deviations from the control group's population mean, and the patient-specific projections of the deviations were used to predict the patients' genetic risks.

In this work, the DL approach is taken for the joint analysis of multi-subject fMRI data consisting of different subject groups. The goal is to estimate the spatial activation maps that are shared across the population as well as the maps that can explain the group differences. The group labels are incorporated through the Fisher discriminant cost added to the DL objective, which encourages clustering of the sparse codes in the same group, while maximizing the distances between group centroids [33]. A second-level analysis is employed here, where each voxel's time-series is regressed first to appropriate reference signals to obtain a single map per subject [36]–[38]. In order to learn the shared and individual brain spatial component maps, two sub-dictionaries are learned that collect the common and the discriminative bases. Then, the vector of coefficients computed per subject capturing the activations of the discriminative component maps is used as the input to the Fisher criterion for group differentiation.

The idea of estimating the shared and discriminative sub-dictionaries was explored for image classification problems in [39], where the set of shared and discriminative features are learned as sub-dictionaries, and the sparse coefficients corresponding to the discriminative sub-dictionaries are used for classification. In the fMRI data analysis, however, it is more natural that the sparsity constraints are imposed on the spatial component maps [40], as it is known that the spatial brain activations are localized and super-Gaussian [41]. These sparse spatial component maps also play the role of the basis set for explaining the data, and the coefficients associated with the discriminative maps are used for group prediction. In a nutshell, the sparsity is imposed on the basis set, rather than the coefficients in our work. Furthermore, instead of employing per-class discriminative sub-dictionaries as in [39], we adopt only a single discriminative sub-dictionary. This is because in fMRI data analysis, even the discriminative component maps are often highly correlated across groups, and estimating per-group maps complicates their interpretation [42]. Moreover, our choice significantly simplifies the parameter selection such as the selection of the model order, and improves the stability of the estimated maps.

For the evaluation of the proposed method, a real task fMRI data set comprising schizophrenic and healthy control subjects is analyzed. To validate the learned brain activation maps, a systematic comparison is done with the maps obtained from the conventional ICA method. To robustify our analysis against local optima, which emerge from solving the nonconvex DL formulation, a rigorous stability-ensuring procedure is employed as well. The estimated maps show that our proposed method can reproduce most of the ICA maps, and also find novel discriminative and interpretable component maps that are not revealed with ICA. A correlation analysis using a set of behavioral test scores further validates the results.

The rest of the paper is organized as follows. The novel

DL formulation is presented in Sec. II. An algorithm to solve the proposed DL problem is derived in Sec. III. The proposed DL method is evaluated and compared with ICA using real task fMRI data in Sec. IV. Finally, the conclusions are provided in Sec. V.

II. Problem Formulation

A. Unsupervised Dictionary Learning

The DL model postulates that the signal vector can be represented by a linear combination of a small number of atoms chosen from a dictionary. Thus, the signals reside in a union of subspaces, and the dictionary constitutes an overcomplete set of basis for the subspaces. DL aims at learning the dictionary from the data.

The conventional DL formulation is an unsupervised learning problem. Let $\mathbf{x}_n \in \mathbb{R}^M$ for $n = 1, 2, \ldots, N$ be the n-th datum. Collect them in the data matrix $\mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$. For a dictionary $\mathbf{D} := [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbb{R}^{M\times K}$, which consists of K atoms $\{\mathbf{d}_k \in \mathbb{R}^M\}$, the DL model assumes that each datum $\mathbf{x}_n \approx \sum_{k=1}^K z_{nk}\mathbf{d}_k = \mathbf{D}\check{\mathbf{z}}_n$, for a sparse coefficient vector $\check{\mathbf{z}}_n = [z_{n1}, \ldots, z_{nk}]^\top$, where $^\top$ denotes the transposition. Upon defining the sparse matrix $\mathbf{Z} := [\check{\mathbf{z}}_1, \ldots, \check{\mathbf{z}}_N] \in \mathbb{R}^{K\times N}$, the DL problem can be formulated as [43]

$$\min_{\mathbf{D},\mathbf{Z}} \ \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 \tag{1a}$$

$$\text{subject to } \mathbf{D} \in \mathcal{D} := \{\mathbf{D} : \|\mathbf{d}_k\|_2 \le 1, \ k = 1, 2, \ldots, K\} \tag{1b}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\|\mathbf{Z}\|_1 := \sum_{n,k} |z_{nk}|$ is the ℓ_1-norm, which promotes sparsity in $\mathbf{Z}$, and $\lambda > 0$ is a parameter that can be varied to adjust the sparsity level of $\mathbf{Z}$. The constraints in (1b) normalize the dictionary atoms, which is necessary due to the scaling ambiguity inherent in the model. That is, scaling the k-th column $\mathbf{d}_k$ of $\mathbf{D}$ by α and the k-th row $\mathbf{z}_k^\top \in \mathbb{R}^{1\times N}$ of $\mathbf{Z}$ by $1/\alpha$ will not alter the product. Thus, the formulation essentially seeks a bi-factorization of data matrix $\mathbf{X}$ into a normalized dictionary matrix $\mathbf{D}$ and a sparse coefficient matrix $\mathbf{Z}$. The DL problem (1) is not a convex optimization problem, but an alternating minimization algorithm can be employed to reach a locally optimal solution [44], [45].

B. Supervised Dictionary Learning

The DL method can also be used for supervised learning tasks. In this case, rather than learning the dictionary to represent the input data with high fidelity, a discriminative dictionary can be learned, which captures the unique traits in the data, characteristic of different classes. In the neuroimaging applications, discriminative DL can reveal neural activity patterns that are unique to different (groups of) subjects.

One way to formulate a supervised DL problem is to augment to the learning objective a classification cost. Similar to [33], we advocate employing the Fisher's discriminant cost [46]. Suppose that the entire data set $\mathbf{X}$ is partitioned to C classes. Let $\mathcal{N}_c$ with cardinality N_c be the set of sample indices belonging to class c, for $c = 1, 2, \ldots, C$. Let $\mathbf{y}_n$ be the feature vector, to be learned by DL, corresponding to the input sample $\mathbf{x}_n$, for all $n = 1, 2, \ldots, N$, and $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. The class mean and the overall mean vectors are defined as

$$\mathbf{m}_c := \frac{1}{N_c}\sum_{n\in\mathcal{N}_c} \mathbf{y}_n \tag{2}$$

$$\mathbf{m} := \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_n \tag{3}$$

respectively. Let us also define so-called the within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$ as

$$\mathbf{S}_w(\mathbf{Y}) := \sum_{c=1}^{C}\sum_{n\in\mathcal{N}_c} (\mathbf{y}_n - \mathbf{m}_c)(\mathbf{y}_n - \mathbf{m}_c)^\top \tag{4}$$

$$\mathbf{S}_b(\mathbf{Y}) := \sum_{c=1}^{C} N_c(\mathbf{m}_c - \mathbf{m})(\mathbf{m}_c - \mathbf{m})^\top \tag{5}$$

respectively. The Fisher criterion aims at learning the features such that they are clustered together in the same class, leading to a small intra-class scatter, and at the same time the class means are far away among others, yielding a large inter-class scatter. Thus, a suitable cost to minimize for the classification task is

$$f(\mathbf{Y}) := \mathrm{tr}\{\mathbf{S}_w(\mathbf{Y})\} - \mathrm{tr}\{\mathbf{S}_b(\mathbf{Y})\} + \|\mathbf{Y}\|_F^2 \tag{6}$$

where the last term ensures the convexity of the cost function with respect to $\mathbf{Y}$ [33], [47].

In fact, upon defining N-by-N matrices $\mathbf{H}_1$ and $\mathbf{H}_2$, where the (n, n')-entry $h_{1,nn'}$ of $\mathbf{H}_1$ is defined as

$$h_{1,nn'} := \begin{cases} \frac{1}{N_c} & \text{if } n, n' \in \mathcal{N}_c \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and all entries of $\mathbf{H}_2$ are equal to $1/N$, the Fisher's objective in (6) can be re-written as

$$f(\mathbf{Y}) = \|\mathbf{Y}(\mathbf{I} - \mathbf{H}_1)\|_F^2 - \|\mathbf{Y}(\mathbf{H}_1 - \mathbf{H}_2)\|_F^2 + \|\mathbf{Y}\|_F^2. \tag{8}$$

Thus, the Hessian $\nabla^2 f(\mathbf{Y})$ is computed as

$$\mathbf{H} := 2\mathbf{I} - 2\mathbf{H}_1 + \mathbf{H}_2 \tag{9}$$

where the symmetry of $\mathbf{H}_1$ and $\mathbf{H}_2$ as well as $\mathbf{H}_2^2 = \mathbf{H}_2 = \mathbf{H}_1\mathbf{H}_2 = \mathbf{H}_2 = \mathbf{H}_1$ are used. It can be easily proved that $\mathbf{H}$ is positive semi-definite by showing the eigenvalues of $\mathbf{H}$ are nonnegative [47]. Furthermore, $f(\mathbf{Y})$ can be simply written as

$$f(\mathbf{Y}) = \mathrm{tr}\{\mathbf{Y}\mathbf{H}\mathbf{Y}^\top\}. \tag{10}$$

Thus, a supervised DL problem can be posed as

$$\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 + \frac{\mu}{2}f(\mathbf{Z}) \tag{11}$$

where $\mathbf{D}$ now captures the discriminative basis for data $\{\mathbf{x}_n\}$, the sparse codes $\{\check{\mathbf{z}}_n\}$ play the role of the features input to the classification cost $f(\cdot)$, and μ is a parameter that balances the reconstructing error and the Fisher cost.

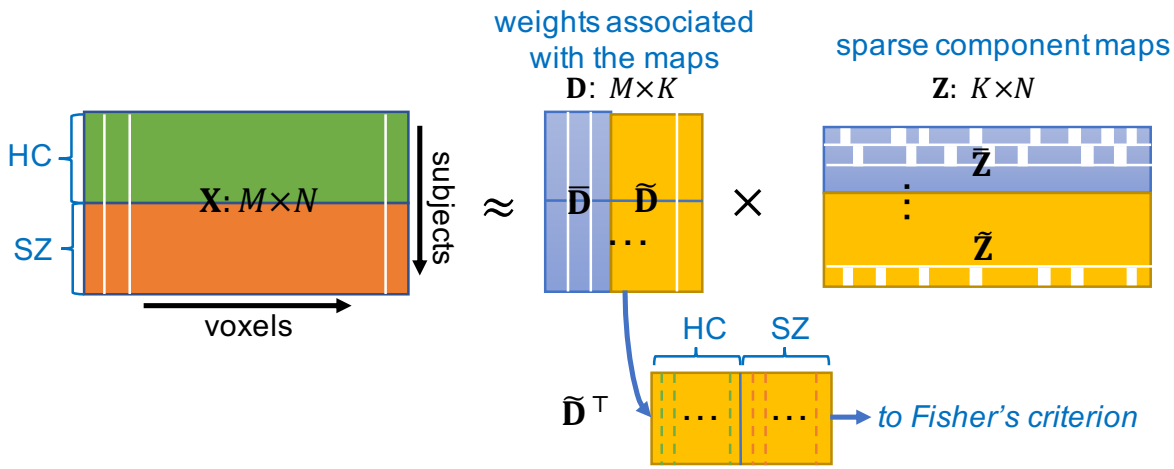


Fig. 1: Illustration of the proposed DL model. The input data $\mathbf{X}$ is a row-wise concatenation of the fMRI volumes of N voxels from M subjects, belonging to HC and SZ groups. The rows of the sparse matrix $\mathbf{Z}$ capture the sparse spatial component maps, where $\overline{\mathbf{Z}}$ collects the common maps and $\widetilde{\mathbf{Z}}$ the discriminative ones. The columns of $\mathbf{D}$ capture subject-wise weights of individual components, i.e., spatial maps. Only the discriminative activations $\widetilde{\mathbf{D}}^\top$ are fed to the Fisher's criterion for group differentiation.

C. Capturing Common and Discriminative Features

In medical image analysis, the focus is not merely on extracting discriminative features for improving classification performance, but rather on obtaining features that can explain both common and individual traits in the groups of samples. For this purpose, a structured dictionary is employed, where $\mathbf{D}$ contains both the common dictionary $\overline{\mathbf{D}}$ that is shared across all class data, and the discriminative dictionary $\widetilde{\mathbf{D}}$, which captures the features reflecting the class differences. That is, let $\mathbf{D} := [\overline{\mathbf{D}}, \widetilde{\mathbf{D}}]$, where $\overline{\mathbf{D}} \in \mathbb{R}^{M\times\overline{K}}$ and $\widetilde{\mathbf{D}} \in \mathbb{R}^{M\times\widetilde{K}}$. Here, $\overline{K}$ is the number of common atoms, and $\widetilde{K}$ the number of discriminative atoms. Likewise, the sparse coefficient matrix $\mathbf{Z}$ is also partitioned compatibly as $\mathbf{Z} = [\overline{\mathbf{Z}}^\top, \widetilde{\mathbf{Z}}^\top]^\top$, where $\overline{\mathbf{Z}} \in \mathbb{R}^{\overline{K}\times N}$ and $\widetilde{\mathbf{Z}} \in \mathbb{R}^{\widetilde{K}\times N}$.

Since only the discriminative features would contribute to predicting the class labels, only $\widetilde{\mathbf{Z}}$ is input to the Fisher cost. The overall supervised DL formulation can thus be constructed as

$$\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 + \frac{\mu}{2} f(\widetilde{\mathbf{Z}}). \qquad (12)$$

Existing methods often adopt per-class discriminative sub-dictionaries and impose additional constraints related to incoherence of the sub-dictionaries, which can improve classification performance in general [33], [34], [39]. Furthermore, to enhance identifiability, a low-rank constraint can be added to the common dictionary for some image classification applications [39]. In this work, our goal is to obtain the brain activation maps from fMRI data that faithfully capture the spatial activation maps and their per-subject contributions. In fMRI data analysis, even the discriminative component maps are often highly correlated across different groups, and estimating per-group maps may hinder their interpretation [42]. Furthermore, proper order selection, which is critical for estimating meaningful maps embedded in noise, becomes more complicated when there are multiple sub-dictionaries. Thus, in this work, we advocate the simple structure involving a single discriminative dictionary along with a common dictionary.

D. Proposed DL Formulation for fMRI Data Analysis

The fMRI data obtained in a scanning session from a subject is a 4-dimensional data consisting of 3-D scans of N voxels taken at T different time points. One can flatten the 3-D volume at each time point into an N-dimensional vector, transforming the data into a T-by-N matrix. To facilitate the processing of multi-subject data sets collected from M subjects, involving M such matrices, the DL model is employed for the second-level analysis. That is, the time dimension in each of the voxel-wise time-series is first regressed away against a reference signal of length T, to obtain a single spatial map per subject, an N-dimensional row vector. More detail is given in Sec. IV-A. Thus, the multi-subject fMRI data $\mathbf{X}$ to be analyzed is a M-by-N matrix, as shown in Fig. 1.

In conventional image classification works, the data vectors belonging to different classes are stacked as *columns* in the data matrix for DL analysis. For example, (12) can be used with $\mathbf{X}^\top$ being the input, in which case, the columns of $\mathbf{D}$ would capture the common and discriminative spatial maps, and the columns of $\mathbf{Z}$ would estimate sparse component activations for each subject. We experimented with the approach in our conference precursor [40]. One issue with this approach is that, for fMRI data, it is more natural to impose sparsity on the component maps, rather than the activations, as it is well-documented that the component maps are super-Gaussian. Existing DL analysis methods for fMRI data also often take this approach [28], [29]. Our own comparison also found that the latter approach tends to obtain more meaningful results [40].

Input: $\mathbf{X}, \mathbf{D}^{(0)}, \mathbf{Z}^{(0)}, \lambda, \mu$
Output: $\mathbf{D}^{(\infty)}$, $\mathbf{Z}^{(\infty)}$

Initialize $\mathbf{D}^{(0)}$ and $\mathbf{Z}^{(0)}$ randomly. Set $\ell = 0$.
While not converged
    Set $i = 0$, $t^{(0)} = 1$, $\mathbf{Z}^{(\ell,i)} = \mathbf{Z}^{(\ell)}$, $\mathbf{W}^{(\ell,i)} = \mathbf{Z}^{(\ell)}$, and $L^{(\ell)} = \lambda_{max}((\mathbf{D}^{(\ell)})^\top \mathbf{D}^{(\ell)})$
    /* *Update* $\mathbf{Z}$ */
    While not converged
        $\mathbf{G}^{(\ell,i)} \leftarrow -(\mathbf{D}^{(\ell)})^\top(\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}^{(\ell,i)})$
        $\mathbf{Z}^{(\ell,i+1)} \leftarrow \mathcal{S}_{\lambda/L}(\mathbf{W}^{(\ell,i)} - \mathbf{G}^{(\ell,i)}/L^{(\ell)})$
        $t^{(i+1)} \leftarrow (1 + \sqrt{1 + 4(t^{(i)})^2})/2$
        $\mathbf{W}^{(\ell,i+1)} \leftarrow \mathbf{Z}^{(\ell,i+1)} + \frac{t^{(i)}-1}{t^{(i+1)}}(\mathbf{Z}^{(\ell,i+1)} - \mathbf{Z}^{(\ell,i)})$
        $i \leftarrow i + 1$
    End While
    Set $s = 0$, $\mathbf{Z}^{(\ell+1)} = \mathbf{Z}^{(\ell,i)}$, and $\mathbf{D}^{(\ell,s)} = \mathbf{D}^{(\ell)}$
    /* *Update* $\mathbf{D}$ */
    Set $\mathbf{A} = \mathbf{Z}^{(\ell+1)}(\mathbf{Z}^{(\ell+1)})^\top$ and $\mathbf{B} = \mathbf{X}(\mathbf{Z}^{(\ell+1)})^\top$
    While not converged
        For $k = 1, 2, \ldots, \overline{K}$
            $\mathbf{u}_k \leftarrow \frac{1}{a_{kk}}(\mathbf{b}_k - \mathbf{D}^{(\ell,s)}\mathbf{a}_k) + \mathbf{d}_k^{(\ell,s)}$
            $\mathbf{d}_k^{(\ell,s+1)} \leftarrow \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k$
            Set the k-th column of $\mathbf{D}^{(\ell,s)}$ to $\mathbf{d}_k^{(\ell,s+1)}$
        End For
        For $k = \overline{K} + 1, \ldots, K$
            $\mathbf{u}_k \leftarrow [a_{kk}\mathbf{I} + \mu\mathbf{H}]^{-1}\left(\mathbf{b}_k - \sum_{k'=1, k' \neq k}^{K} a_{k,k'}\mathbf{d}_{k'}^{(\ell,s)}\right)$
            $\mathbf{d}_k^{(\ell,s+1)} \leftarrow \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k$
            Set the k-th column of $\mathbf{D}^{(\ell,s)}$ to $\mathbf{d}_k^{(\ell,s+1)}$
        End For
        $s \leftarrow s + 1$
    End While
    Set $\mathbf{D}^{(\ell+1)} = \mathbf{D}^{(\ell,s)}$
    $\ell \leftarrow \ell + 1$
End While
$\mathbf{D}^{(\infty)} \leftarrow \mathbf{D}^{(\ell)}$ and $\mathbf{Z}^{(\infty)} \leftarrow \mathbf{Z}^{(\ell)}$

TABLE I: Algorithm for solving (13).

In this work, we focus on stacking the per-subject maps as rows in $\mathbf{X}$, which is then input to DL. Thus, the rows of $\mathbf{Z}$ correspond to the common and discriminative spatial component maps. A row in $\mathbf{D}$ is the corresponding subject's weights associated with the component maps. A column in $\mathbf{D}$ can be regarded as the weights associated with the corresponding map, where each entry represents each subject's contribution. Since only the discriminative activation coefficients $\widetilde{\mathbf{D}}$ would contribute to label prediction, $\widetilde{\mathbf{D}}^\top$ is input to the Fisher's criterion. This is summarized in Fig. 1, where the subjects are shown to be partitioned into two classes, namely the healthy control (HC) group and the schizophrenic (SZ) group. The group labels serve as the supervision signals. Thus, our DL formulation for fMRI data analysis is finally given by

$$\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 + \frac{\mu}{2}f(\widetilde{\mathbf{D}}^\top). \tag{13}$$

III. Algorithm Derivation

The proposed formulation (13) is a nonconvex optimization problem, and thus it is difficult to obtain a globally optimal solution efficiently. On the other hand, it is observed that when $\mathbf{D}$ is fixed, the optimization with respect to (w.r.t.) $\mathbf{Z}$ is essentially a convex least absolute shrinkage and selection operator (LASSO) problem. Likewise, when $\mathbf{Z}$ is fixed, the problem for $\mathbf{D}$ is also convex as f is convex w.r.t. $\widetilde{\mathbf{D}}$, the squared error term is convex, and set $\mathcal{D}$ is convex as well. Thus, the block coordinate descent (BCD) method can be employed by alternating between blocks $\mathbf{Z}$ and $\mathbf{D}$, which converges to a locally optimal solution.

The update for $\mathbf{Z}$ when $\mathbf{D}$ is fixed to its ℓ-th iterate $\mathbf{D}^{(\ell)}$ is to solve

$$\mathbf{Z}^{(\ell+1)} := \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1. \tag{14}$$

This problem is an instance of the well-known LASSO

problem applied to each column of $\mathbf{X}$. There are a variety of methods to solve this problem. In this work, we adopt the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [48], which iteratively solves a proximal problem involving a linear approximation of the differentiable part of the objective. The next iterate is chosen based on the "optimal" first-order method. The FISTA thus requires the gradient of the differentiable part $h(\mathbf{Z};\mathbf{D}^{(\ell)}) := \frac{1}{2}\|\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}\|_F^2$ and the Lipschitz constant of the gradient, which are given by

$$\frac{\partial h(\mathbf{Z};\mathbf{D}^{(\ell)})}{\partial \mathbf{Z}} = -(\mathbf{D}^{(\ell)})^\top(\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}) \tag{15}$$

$$L^{(\ell)} = \lambda_{\max}((\mathbf{D}^{(\ell)})^\top\mathbf{D}^{(\ell)}) \tag{16}$$

respectively, where $\lambda_{\max}(\mathbf{M})$ denotes the largest eigenvalue of a matrix $\mathbf{M}$. The FISTA update is described in lines 4–10 in Table I, where the (i,j)-entry of $\mathcal{S}_\alpha(\mathbf{M})$ is given by soft-thresholding the (i,j)-entry m_{ij} of $\mathbf{M}$, that is, $[\mathcal{S}_\alpha(\mathbf{M})]_{ij} := \mathrm{sign}(m_{ij}) \cdot \max\{0, |m_{ij}| - \alpha\}$.

When $\mathbf{Z}$ is fixed at $\mathbf{Z}^{(\ell+1)}$, the update for $\mathbf{D}$ is done by solving [cf. (10)]

$$\mathbf{D}^{(\ell+1)} := \arg\min_{\mathbf{D}\in\mathcal{D}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}^{(\ell+1)}\|_F^2 + \frac{\mu}{2}\mathrm{tr}\left\{\widetilde{\mathbf{D}}^\top\mathbf{H}\widetilde{\mathbf{D}}\right\}. \tag{17}$$

As the constraint $\mathbf{D} \in \mathcal{D}$ is decoupled to individual columns $\{\mathbf{d}_k\}$ of $\mathbf{D}$, another layer of BCD can be employed to solve (17) [44]. First, upon defining $\mathbf{A} := \mathbf{Z}^{(\ell+1)}(\mathbf{Z}^{(\ell+1)})^\top$ and $\mathbf{B} := \mathbf{X}(\mathbf{Z}^{(\ell+1)})^\top$, it is noted that $\mathbf{D}^{(\ell+1)}$ in (17) can be obtained from the equivalent formulation given by

$$\mathbf{D}^{(\ell+1)} = \arg\min_{\mathbf{D}\in\mathcal{D}} \frac{1}{2}\mathrm{tr}\{\mathbf{A}\mathbf{D}^\top\mathbf{D}\} - \mathrm{tr}\{\mathbf{B}\mathbf{D}^\top\} + \frac{\mu}{2}\mathrm{tr}\{\mathbf{H}\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}^\top\}. \tag{18}$$

Thus, in iteration s, updating the k-th column of $\mathbf{D}$ with all other columns fixed at $\mathbf{d}_{k'}^{(\ell)} = \mathbf{d}_{k'}^{(\ell,s)}$ for $k' = 1,\ldots,k-1,k+1,\ldots,K$, needs to be done differently depending on whether the k-th atom belongs to the common dictionary $\overline{\mathbf{D}}$ or the discriminative dictionary $\widetilde{\mathbf{D}}$. That is, for $k = 1,2,\ldots,\overline{K}$, the following steps are used to update $\mathbf{d}_k$.

$$\mathbf{u}_k = \frac{1}{a_{kk}}\left[\mathbf{b}_k - \sum_{k'\neq k} a_{k,k'}\mathbf{d}_{k'}^{(\ell,s)}\right] \tag{19}$$

$$\mathbf{d}_k^{(\ell,s+1)} = \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k \tag{20}$$

where $a_{kk'}$ is the (k,k')-entry of $\mathbf{A}$, and $\mathbf{a}_k$ and $\mathbf{b}_k$ are the k-th columns of $\mathbf{A}$ and $\mathbf{B}$, respectively. Similarly, for $k = \overline{K}+1,\ldots,K$,

$$\mathbf{u}_k = (a_{kk}\mathbf{I} + \mu\mathbf{H})^{-1}\left[\mathbf{b}_k - \sum_{k'\neq k} a_{k,k'}\mathbf{d}_{k'}^{(\ell,s)}\right] \tag{21}$$

$$\mathbf{d}_k^{(\ell,s+1)} = \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k. \tag{22}$$

The updates (19)–(22) are repeated over s until convergence, and the converged $\mathbf{D}^{(\ell,s)}$ is the optimal solution to (18). The overall dictionary update steps correspond to lines 12–25 in Table I. The entire algorithm for solving (13) is provided in Table I.

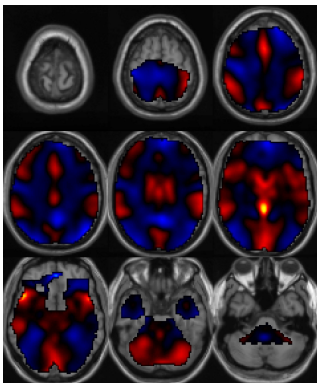
A remark on the computational complexity of the algorithm is in order. For updating $\mathbf{Z}$ in lines 4-10, $(\mathbf{D}^{(\ell)})^\top\mathbf{X}$ and $(\mathbf{D}^{(\ell)})^\top\mathbf{D}^{(\ell)}$ can be computed outside the while loop. Thus, the dominant operation is the multiplication of $(\mathbf{D}^{(\ell)})^\top\mathbf{D}^{(\ell)}$ with $\mathbf{Z}^{(\ell,i)}$, which can be done in $O(K^2N)$ operations. For updating $\mathbf{D}$, the most demanding operation is the inversion of an M-by-M matrix in line 20, which requires $O(M^3)$ operations. Since this has to be done $\widetilde{K}$ times, the computational cost per while iteration is $O(M^3K)$ (assuming that K and $\widetilde{K}$ are similar.) The overall computational cost is also dependent on the number of iterations for the while loops. In the computational setup to be explained in Sec. IV, the average number of while iterations for the $\mathbf{Z}$-update was around 200, and the same for the $\mathbf{D}$-update was around 30. Finally, the outer while loop typically took around 150 iterations. Executing the algorithm on an Intel Zeon processor with 8 cores at 3.4 GHz clock speed, typically took 5 to 30 minutes.

IV. Evaluation

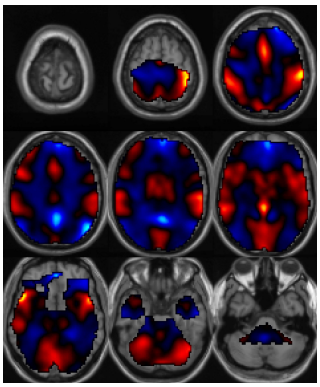
A. FMRI Data Set and Regression Analysis

The proposed method was evaluated using real fMRI data. We used task fMRI data sets published by MIND Clinical Imaging Consortium (MCIC), which can be accessed publicly at https://coins.trendscenter.org. The data set contains 150 healthy control (HC) subjects and 121 patients with schizophrenia (SZ), collected as the subjects, who performed the auditory oddball (AOD) task. The subjects were presented with three types of auditory stimuli: the standard, novel, and target stimuli. The standard stimulus was 1 kHz-tones that occurred with probability 0.82. The novel stimulus was randomly generated, complex and non-repeating digital noise presented with probability 0.09. The target stimulus was 1.2 kHz-tones that occurred with probability 0.09. The subjects were instructed to press a button with their right index fingers when they heard the target tones. For each run, overall 90 stimuli with a duration of 200 ms were played at random intervals. The stimulus sequences were designed to produce orthogonal BOLD responses [49]. Furthermore, the order of the novel and the target stimuli was shuffled between runs to make sure the responses do not depend on the stimulus order. More detailed description of the data sets can be found in [50]. The preprocessing steps of individual fMRI data including motion correction, artifact removal, spatial normalization, smoothing and subsampling are explained in [38] and [50].

The regressors were then created by modeling the target and standard stimuli as the convolution of the delta functions capturing the stimulus onsets and the default SPM HRF in addition to their temporal derivatives [51]. Finally, the voxel-wise time-series was regressed in the GLM framework and the contrast images between the target versus the standard stimuli were obtained as the input matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$ for the proposed method, where $M = 271$ subjects and



(a) HC group



(b) SZ group

Fig. 2: Averages of the rows of $\mathbf{X}$ for the HC and the SZ groups

$N = 48,546$ voxels. More details on feature extraction can be found in [52]. The contrast images convey the difference in spatial activations associated with the detection of the target tone and pressing the button as instructed, relative to the background state, which involves listening the standard and novel tones, which require no response. The task involves a variety of cognitive processes as well as auditory and motor functions. Fig. 2 shows the averages of the rows of X for the HC and the SZ groups. While there are common activations across the groups, groups differences are also clearly visible. Note that in order to increase the likelihood of getting components in the areas of interest, non-brain voxels were removed through masking and the number of voxels was reduced to N.

B. Parameter Tuning

The model presented in Sec. II contains some model parameters that must be determined. These are the size the of the shared dictionary $\overline{K}$, the size of the discriminative dictionary $\widetilde{K}$, the parameter λ for adjusting sparsity of $\mathbf{Z}$, and the weight μ for the Fisher cost. We employed cross-validation to fix these parameters.

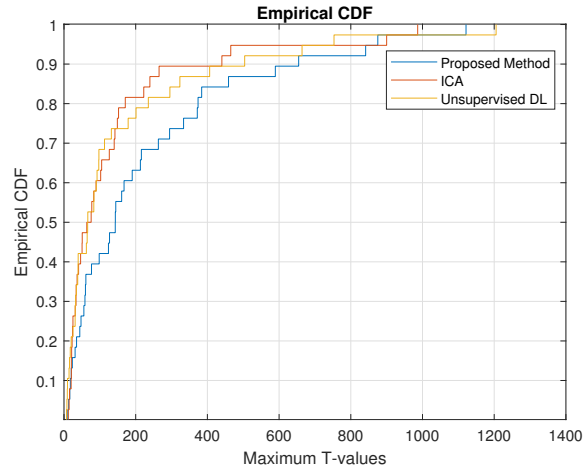


Fig. 3: The CDFs of the maximum T-map values from the proposed algorithm, unsupervised DL in [28] and the ICA method.

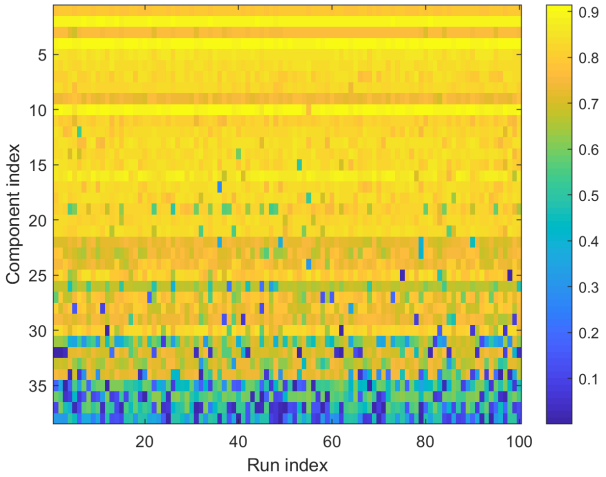
First, a training set and a validation set were created by randomly selecting subjects from the entire data sets. As the numbers of the HC and SZ subjects are not the same (150 and 121, respectively), a balanced training set was obtained by randomly choosing 85 HC subjects and 85 SZ subjects. Similarly, 36 HC and 36 SZ subjects were picked for the validation set. A total of 100 different training/validation sets were constructed in this way. Then a grid search was performed over $\overline{K}$, $\widetilde{K}$, λ and μ to find the combination that yielded the best average classification accuracy on the validation sets, averaged over the 100 sets.

The classification was done as follows. Based on the spatial maps $\mathbf{Z}$ obtained from the training set, the weights associated with corresponding maps $\mathbf{D} = [\overline{\mathbf{D}}, \widetilde{\mathbf{D}}]$ were computed by the least-square (LS) regression for the validation set. Then, for each subject m, the m-th row of $\widetilde{\mathbf{D}}$ was input to the 1-nearest neighbor classifier. The best average classification accuracy achieved was 68% with parameters $\overline{K} = 12$, $\widetilde{K} = 26$, $\lambda = 0.003$, and $\mu = 0.2$. This accuracy is slightly higher than what was reported in a prior study using the same data set [52], which was around 66%, although as in [52] we did not try to optimize the classifier itself.

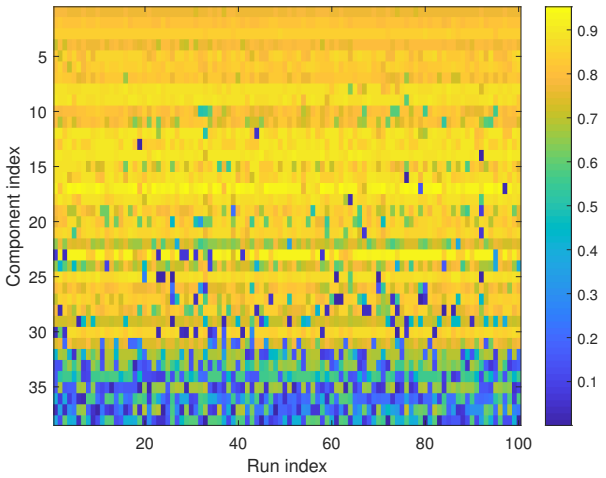
C. Obtaining Stable Maps

As was discussed in Sec. II-D, (13) is a nonconvex optimization problem, and thus only locally optimal solutions can be obtained in practice. Therefore, the algorithm in Table I would yield different solutions and thus different spatial maps, depending on the initialization. Consequently, it is prudent to use the spatial maps that are more consistently obtained across multiple runs using random initializations in order to improve the reliability of the conclusions drawn from the analysis.

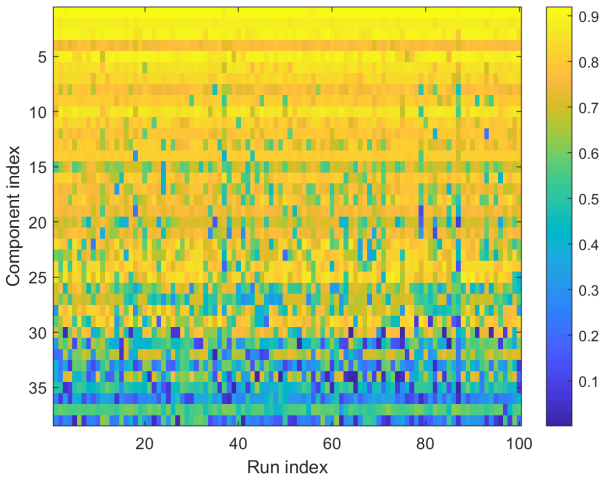
In this work, the framework proposed in [53] was adapted for our setting. First, it is noted that permuting the columns of $\mathbf{D}$ and the rows of $\mathbf{Z}$ using the same permutation does not alter the objective function value. Thus, this permutation



(a) Proposed method



(b) Unsupervised DL



(c) ICA

Fig. 4: Pearson correlation between the component maps from each run and the T-maps.

ambiguity must be resolved in order to compare any two sets of maps. This can be done by solving a linear assignment problem (LAP), which finds the bipartite matching that minimizes the total edge weight for a weighted bipartite graph [54]. Specifically, suppose that one obtained K spatial maps from each of the R runs and denote the k-th map from the r-th run as $\mathbf{z}_k^{(r)}$, where $k = 1, 2, \ldots, K$ and $r = 1, 2, \ldots, R$. The weight between the k-th map $\mathbf{z}_k^{(r)}$ from run r and the k'-th map $\mathbf{z}_{k'}^{(r')}$ from run r' is defined as $w^{(r,r')}(k,k') := 1 - |(\mathbf{z}_k^{(r)})^\top \mathbf{z}_{k'}^{(r')}| / \|\mathbf{z}_k^{(r)}\|_2 \|\mathbf{z}_{k'}^{(r')}\|_2$. The LAP is to find a bijection $f : \{1, \ldots, K\} \to \{1, \ldots, K\}$ to assign each map k in run r to map $k' = f(k)$ in run r' such that $\sum_{k=1}^{K} w^{(r,r')}(k, f(k))$ is minimized. Let us denote the resulting minimum cost as $w^{(r,r')*}$. The problem can be solved using Hungarian algorithm [55].

Then, a weighted graph is constructed where the R runs constitute the nodes, and the edge weight between nodes r and r' is $w^{(r,r')*}$. A minimum spanning tree (MST) is found on this graph, which is defined as the subgraph connecting all nodes in the graph with minimum total weights [56]. Then, the node with the maximum number of edges is selected as the central node. If there are multiple nodes that have the same maximum number of edges, then the one with the minimum total edge weight is chosen. The central node is essentially the run that yielded the spatial maps that are highly correlated with the most other runs. Subsequently, the maps in all runs are reordered so that they are aligned with the maps in the central node, to mitigate the permutation ambiguity.

After the alignment, the one-sample t-test statistic is computed for each voxel by regarding R runs as R samples. The resulting T-map captures the individual voxel's stability in each map. Then, the run r^* whose spatial maps have the highest correlation with this T-map is found. The maps from run r^* are our most stable maps.

We applied the aforementioned strategy for analyzing the data set $\mathbf{X}$ containing 121 HC and 121 SZ subjects. This data set $\mathbf{X}$ is the concatenation of the training and the validation sets that yielded the best classification accuracy during the parameter tuning stage explained in Sec. IV-B. We ran the algorithm in Table I with $R = 100$ random initializations of $\mathbf{D}^{(0)}$ and $\mathbf{Z}^{(0)}$. For comparison, we also ran the unsupervised DL method proposed in [28] as well as the entropy bound minimization (EBM)-based ICA algorithm in [57] with R random initializations, and applied the same strategy.

To compare the stability of the proposed algorithm with unsupervised DL method and the ICA algorithm, the maximum value of the T-map (maximized over the voxels) was computed for each map. Although the spatial distribution of the T-values is of course not uniform, having only a few voxels estimated stably while others unstably is unlikely. Thus, the maximum T-values for each component is a reasonable indicator of how stable the entire component is. In Fig. 3, the empirical cumulative distribution functions (CDFs) of the resulting maximum T-values are plotted. The number of components was set to $K = \overline{K} + \tilde{K} = 38$ for all the methods. It can be seen that a given percentile of the

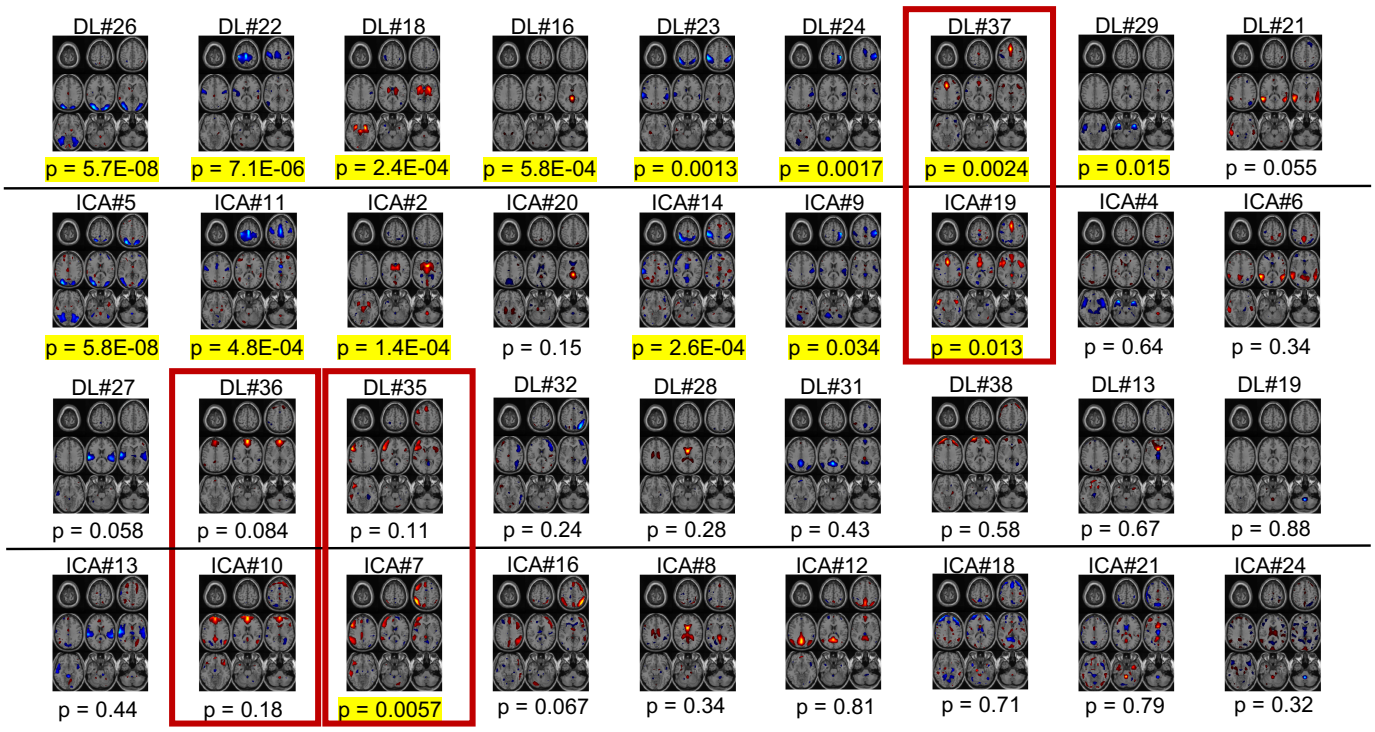


Fig. 5: The discriminative maps $\widetilde{\mathbf{Z}}$ from DL and the matching ICA maps. The p-values lower than 0.05 are highlighted in yellow. The ICA maps in the red boxes were found to be split into multiple DL maps (see Fig. 7). It can be seen that the DL maps are not only readily interpretable, but also more focal and group-different with higher significance than the ICA counterparts in many cases.

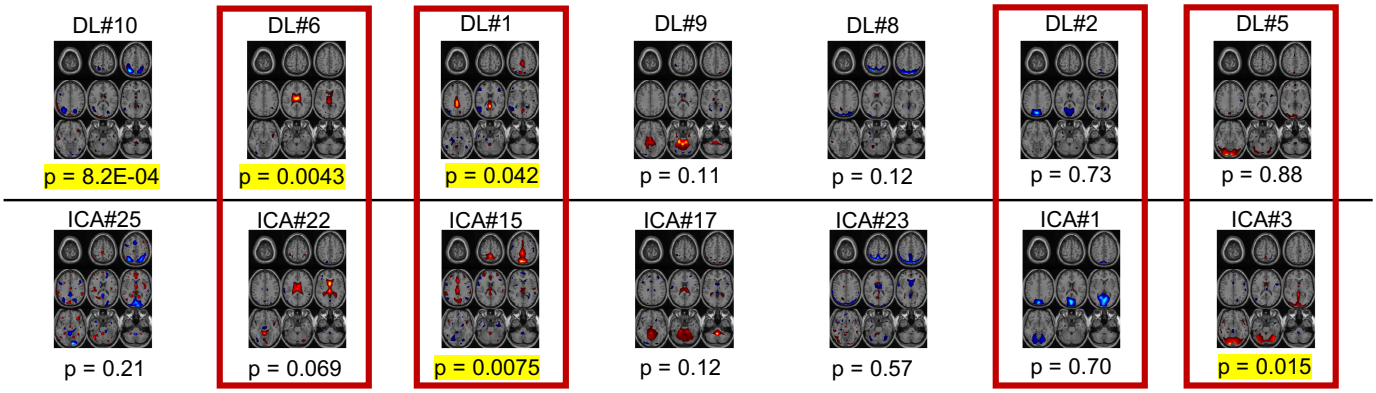


Fig. 6: The common maps $\overline{\mathbf{Z}}$ from DL with matching ICA maps. The DL maps are again cleaner than the ICA maps. Some common maps turn out to be group-different, which may be due to the local optima of the proposed optimization formulation.

maximum T-value is higher in the proposed method than the unsupervised DL or the ICA algorithms, which means that the proposed method provides maps that are more stable. In particular, it is noted that our method yields maps that are more stable than those from unsupervised DL, indicating that exploiting the label information helps stability.

Fig. 4 shows the actual correlation values of the maps obtained from the individual runs with the T-maps. Fig. 4a corresponds to the proposed method, Fig. 4b the unsupervised DL, and Fig. 4c the ICA. It is clearly seen that the proposed method tends to yield more consistent maps than the unsupervised DL and ICA counterparts. Thus, the maps obtained from the proposed method are sufficiently stable for further analysis.

Remark: It should be noted that the ICA and DL methods are inherently different and with a simpler but less flexible ICA algorithm such as Infomax [58], one would obtain more stable components, but at the expense of limited approximation to the density of the underlying sources. When more flexible and powerful algorithms like EBM are used, a common practice is to use a scheme for selecting the most consistent one among multiple runs [31], [53].

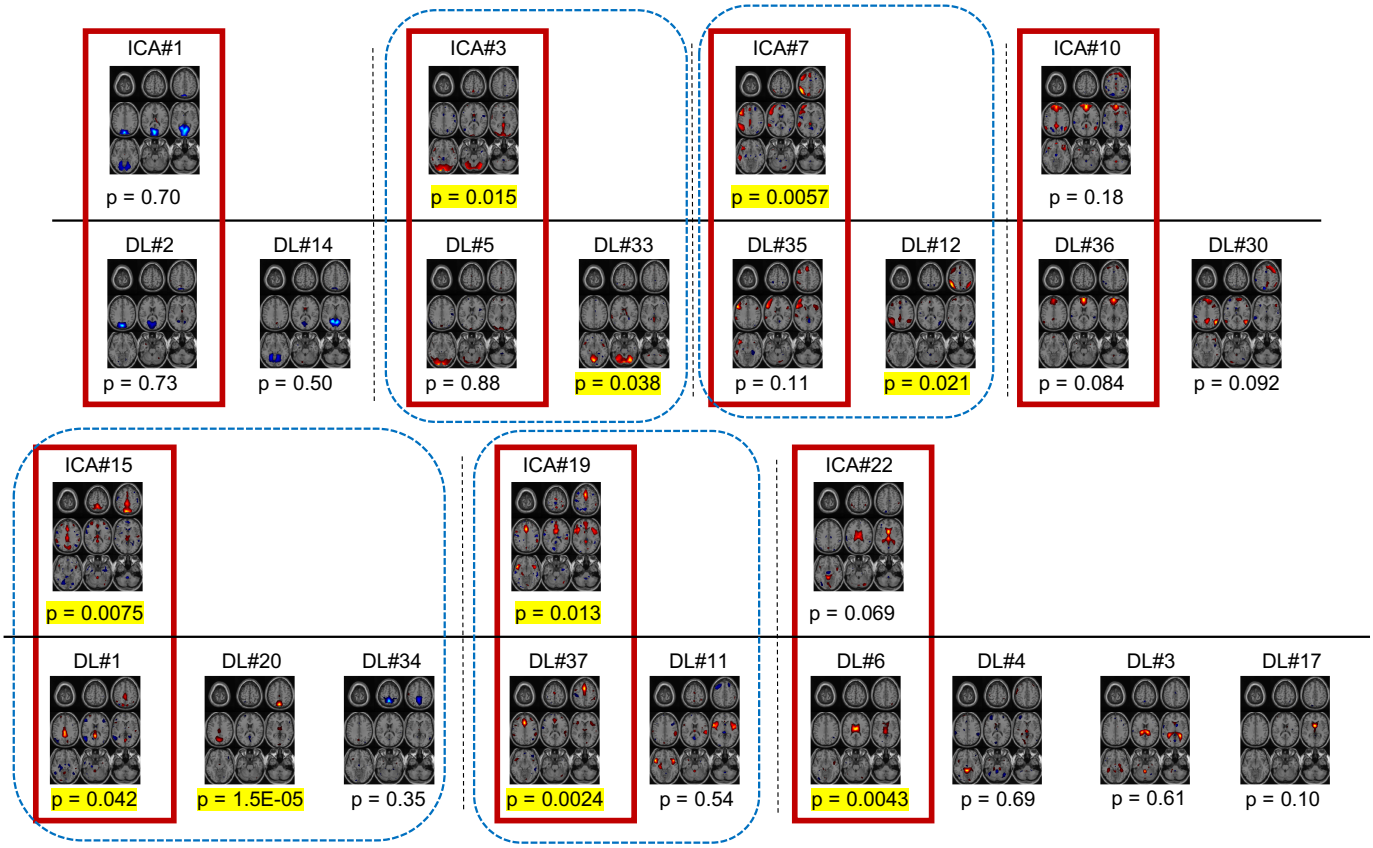


Fig. 7: The ICA maps that split into multiple DL maps. The red boxes indicate the matching components between DL and ICA, found by solving the LAP. In the blue dashed boxes, group-different ICA components are seen to be split into multiple DL components that include the ones that are *not* group-different, highlighting that DL method is finding more localized discriminative regions without losing common features.

D. Analysis of Obtained Spatial Maps

The common and discriminative brain activation maps obtained from Sec. IV-C are analyzed in this section. It is recalled that the k-th column $\mathbf{d}_k$ of the learned dictionary $\mathbf{D}$ can be interpreted as the weights associated with the k-th component map, which is the k-th row $\mathbf{z}_k$ of $\mathbf{Z}$. Thus, the group difference in the activation can be revealed by performing a two-sample t-test on $\mathbf{d}_k$, whose test statistic is defined as

$$t_k = \frac{\hat{\mu}_{HC,k} - \hat{\mu}_{SZ,k}}{\sqrt{\frac{\hat{\sigma}^2_{HC,k}}{M_{HC}} + \frac{\hat{\sigma}^2_{SZ,k}}{M_{SZ}}}} \tag{23}$$

where M_{HC} and M_{SZ} are the numbers of the HC and SZ subjects, respectively, which are both equal to 121 in the data set used, and $\hat{\mu}_{HC,k}$ and $\hat{\mu}_{SZ,k}$ are the sample means of the entries of $\mathbf{d}_k$ corresponding to the HC subjects and the SZ subjects, respectively. Sample variances $\hat{\sigma}^2_{HC,k}$ and $\hat{\sigma}^2_{SZ,k}$ are defined in the same way. To determine components with significance, we have used a p-value of 0.05 for the statistic (uncorrected).

In Figs. 5–8, the learned spatial maps are plotted. All the spatial map plots in this work represent thresholded Z-maps, where the Z-values are calculated per voxel by dividing the entries in $\mathbf{z}_k$ by the standard deviation of the entries, and then the Z-values whose absolute values are larger than 2 are plotted. Furthermore, as there is a sign ambiguity on the signs of a component map $\mathbf{z}_k$ and its activations $\mathbf{d}_k$, the sign is fixed such that t_k in (23) is nonnegative. That is, all the component maps shown have higher activations in the HC group than the SZ group. The red colors in the map represent positive values in the Z-map, with bright yellow indicating the highest intensity, and the blue colors represent negative values, with bright skyblue indicating the highest intensity.

To assess the performance of the proposed approach, we compared the maps with those obtained from ICA, another data-driven method, and one that has been now well established for such studies[1]. We reduced the dimensionality prior to ICA analysis using an information theoretic order selection method based on MDL, which also takes sample

[1]The shared and specific independent component analysis (SSICA) algorithm proposed in [59] can be used to find the shared and the group-specific components based on structural constraints on the mixing matrix. However, the algorithm finds a separate set of components for each group, whereas our method finds one set for the discriminative components of all groups. Thus, matching and comparing the resulting maps with our DL component maps would entail additional steps. Furthermore, SSICA is originally designed for the first-level analysis, and thus would require additional investigation to adapt for the second-level analysis done in the present work. For these reasons, the well-established ICA method is adopted for comparison with our DL method.

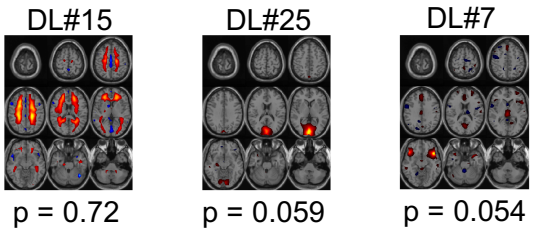


Fig. 8: DL maps without matching ICA maps.

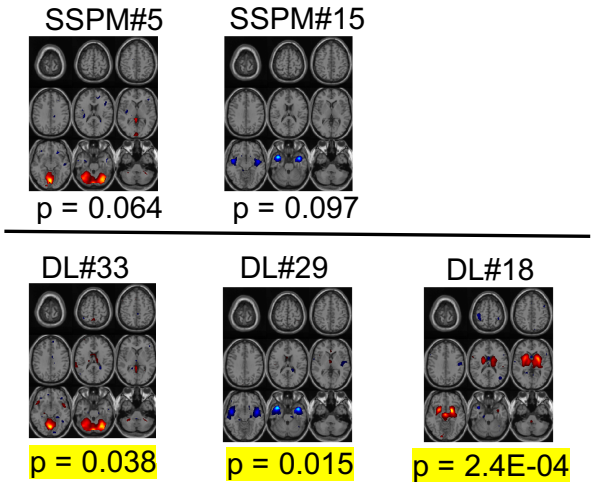


Fig. 9: Samples of the maps from the sparse SPM and the corresponding maps from the proposed DL method. Note that DL#18 was not identified from the sparse SPM

dependence into account [60]. The order for ICA was determined to be 25, thus enabling better generalization performance for ICA. That is, the ICA maps showed the best statistical significance and interpretability around order 25, while deviating from this order resulted in the maps with significantly degraded p-values and interpretability. Then, stable maps were determined as explained in Sec. IV-C. For this, the rectangular LAP was solved to match the $K = 38$ component maps from the DL and $K = 25$ maps from the ICA. In addition, the component maps were also further checked by visual inspection to make sure the results were reliable.

Figs. 5 and 6 show the maps learned from DL and ICA that are found matching both from LAP and visual inspection. Fig. 5 lists the discriminative maps $\widetilde{\mathbf{Z}}$ from the DL and the matching ICA maps, whereas Fig. 6 depicts the common maps $\overline{\mathbf{Z}}$ and those matching from ICA. The maps in the top row of each figure are the maps from the DL and the bottom from the ICA. The map index k is indicated on the top of each map. The indices $k = 1, 2, \ldots, \overline{K} = 12$ correspond to common components, and $k = \overline{K}+1 = 13, \ldots, \overline{K}+\widetilde{K} = 38$ to discriminative components. The p-values are provided at the bottom of the maps. When the map is significant ($p < 0.05$), the p-value is highlighted in yellow. It can be seen that DL identifies 25 component maps that are in good match with the ICA-based maps, which validates the map estimation performance of the DL algorithm[2]. Furthermore, it is observed from Fig. 5 that many of the maps in $\widetilde{\mathbf{Z}}$ indeed show significant group difference. In fact, the p-values of the discriminative maps turn out to be *smaller* than the p-values of the corresponding ICA maps in most cases, which verifies that the supervised DL formulation is working as designed. In particular, DL#24 and DL#22, which can be ascribed to motor and sensory motor functions, respectively, and thus are strongly related with the AOD task, are showing more significant group difference than their ICA counterparts ICA#9 and ICA#11. Furthermore, the DL maps #16 and #29 are interesting, since the maps are group-different while the matching ICA maps #20 and #4 are not. The map DL#29 is showing significant activations in inferior and middle temporal gyrus, which are known to be associated with language and semantic memory processing, visual and facial perception, and multimodal sensory integration. On the other hand, the map DL#16 is showing activations in the brain-stem area, directly related to controlling respiration, pain modulation, motor, and cardiac output [61]. These highlight that the proposed DL is finding more discriminative maps by incorporating the label information. Many maps in Figs. 5 and 6 are readily interpretable as motor (DL#23, DL#24), sensory motor (DL#22), auditory (DL#27), anterior default mode network (DMN) (DL#36), posterior DMN (DL#31) and frontal parietal regions (DL#35). It can be observed that the DL-based maps are often much more localized and cleaner than the ICA counterparts. This is because the DL is formulated to estimate sparse component maps.

As the DL aims at finding sparse maps, we observe splitting of components into multiple sparse maps [62], when compared with maps estimated using ICA. This is observed in Fig. 7, which depicts some of the ICA maps containing activation regions explained by multiple DL maps. The red boxes in Fig. 7 indicate the matching found by LAP, which are also shown in Figs. 5–6. In particular, it is observed that the ICA maps in the blue boxes are group different, which are split into a set of DL maps that include the maps that are *not* group different. For example, component in DL#12 is showing activation in frontal parietal region associated with attention network, while the matching ICA map #7 has activations in other areas not significant and not related to the task. This underlines that DL map is finding more localized regions that are group different, without losing the common regions, based on our DL formulation. Similar trends can be found in other components shown in Fig. 7. DL map #37 is showing activations in anterior cingulate cortex with a lower p-value, while ICA#19 has activations in insular areas as well. DL map #20 is related to parietal areas and showing very significant activations compared to ICA#15.

[2]As was explained, the colors in the maps capture the signs of the map values. Since we fixed the sign ambiguity for each map such that the two-sample t-test statistic is positive, the color may be misleading when the t-statistic is not significant. Since the p-values for the three DL maps #31, #32 and #38 are quite high, we decided to neglect the signs for these maps and match them with the ICA maps #12, #16, and #18, respectively, in spite of the color mismatch.

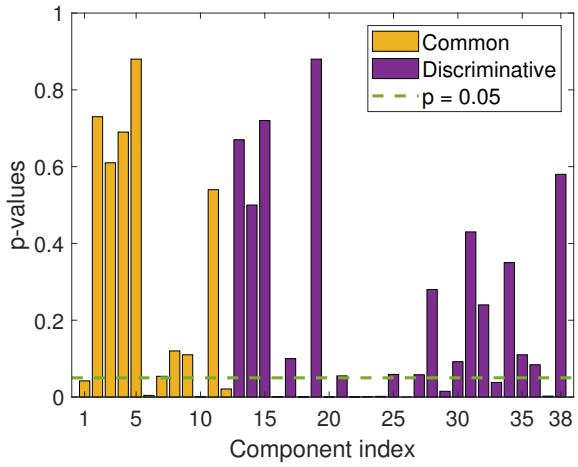


Fig. 10: Bar plot of the p-values of all DL maps.

Fig. 8 shows the DL maps that are not matched with any of the ICA maps. These are the maps that were not assigned from solving the rectangular LAP, nor were associated with the split maps in Fig. 7. Although the p-values for components DL#25 and DL#7 are not lower than the hard threshold 0.05, they are quite close to it, potentially contributing meaningfully for discrimination. Furthermore, they correlate well with known functional areas, such as the visual (DL#25) and the insular (DL#7) areas. One component (DL#15) is picking up the white matter area, presumably owing to noise and artifacts present in the data. Note that there are no ICA maps left unmatched with DL maps.

Remark: Note that even though our proposed formulation encourages the discriminative maps to be collected separately from the common maps, the formulation is nonconvex, hence is prone to have local optima. Thus, even after our effort to ensure stability, the found common maps may turn out to be discriminative, and the found discriminative maps common. However, the general tendency can be seen to be clearly toward the intended separation. From Fig. 10, it can be seen that most of the p-values for the common maps tend to be much larger than the 0.05 threshold, while for the discriminative maps, the p-values are often quite small. In fact, the five smallest p-values are all associated with the discriminative maps. Some summary statistics derived from the p-values can help convey the point more clearly. The geometric mean of the p-values of the common maps is computed to be 0.091, whereas the same for the discriminative maps is equal to 0.017, indicating that the discriminative maps tend to be more significant in detecting the group differences. Similarly, geometric median (i.e., $\exp(\text{med}\{\log p_i\})$) for the common maps is equal to 0.11 while the same for the discriminative maps is 0.070, again indicating that the discriminative maps are more significantly group different.

We also compared the spatial maps from our DL method with the maps obtained from the sparse SPM algorithm in [28], which is an unsupervised DL method. The p-values of the maps were computed again using the two-sample t-tests. The comparison between the two sets of maps reveals that most of the maps from the two methods match very well. However, it is also observed that some of the maps identified by our proposed DL method show group differences, whereas the corresponding components from the sparse SPM are not. This can be ascribed to the use of the label information in our method. The maps are shown in Fig. 9. Note that DL#18 was not identified from the sparse SPM, but it actually exhibits highly significant group-difference. On the average, the geometric mean of the p-values of the maps from the sparse SPM turns out be 0.036, whereas the same for the maps due to our method is 0.028. (If only the discriminative maps $\widetilde{\mathbf{Z}}$ is considered, the number goes down to 0.017.) In summary, our method extracts faithfully the components of the sparse SPM but with enhanced group-difference in the discriminative maps.

In the conference paper [40], we also performed a comparison with the method in [39], which is a discriminative DL method, developed for image classification. Specifically, formulation (P1) in [40] contains a low-rank constraint on the common dictionary, and the sparse coefficients corresponding to the discriminative dictionary are used for classification. It was observed that the resulting spatial maps showed less similarity to the ICA maps, than our proposed method. It was also found that the maps from our method (equivalently, formulation (P2) in [40]) were often more interpretable and showed more significant group differences. The detailed estimated maps and discussion can be found in [40].

In conclusion, the proposed algorithm can find many component maps that can be readily cross-validated with ICA. By exploiting sparsity, the maps from DL are cleaner and more localized. Through incorporation of the label information during the learning process, our method tends to discover more meaningful activations, and in particular those that provide more specific areas with greater sensitivity in differentiating the patients from the healthy controls.

E. Correlation Analysis with Behavioral Variables

The spatial component maps obtained from the DL analysis can be further validated and interpreted by using the behavioral test score data that we have. A total of 105 behavioral variables (BVs) were available for 193 of the 271 subjects. Out of all the BVs, 7 were found to be significantly group different based on a two-sample t-test. Table II lists the 7 BVs along with their p-values.

Then, we computed the Pearson correlation between the BVs and the weight vector associated with k-th component map $\mathbf{d}_k$ (only the part for the 193 subjects for whom the BVs are available) learned from the DL analysis. Table III lists all the component maps that are significantly correlated with each of the BVs, that is, the components that yield p-values less than 0.05 from one-sample t-tests on the Pearson correlations.

The maps listed in Table III are sorted in the order of descending correlation. It can be observed that the correlated

BV index	#9	#23	#32	#37	#53	#88	#94
Behavioral test	BVRT	FAS	WMS-3: Logical Memory I	TMT	WMS-3: Logical Memory II - Delay	WAIS-3	HVLT: Immediate
p-value	5.6E-13	3.3E-12	0.0E+00	1.1E-05	0.0E+00	7.8E-06	0.0E+00

TABLE II: Behavioral tests and corresponding p-values.

BV index	DL spatial map index
#9	DL#16[p=2E-6], DL#37, DL#22, DL#20, DL#26, DL#25, DL#10, DL#19, DL#32, DL#14 [p=2E-2]
#23	DL#26 [7E-4], DL#16, DL#37, DL#23 [2E-2]
#32	DL#16 [3E-4], DL#22, DL#20, DL#10, DL#25, DL#24, DL#19, DL#6, DL#23, DL#12, DL#37, DL#26 [4E-2]
#37	DL#16 [3E-5], DL#10, DL#22, DL#21, DL#20, DL#14 [4E-2]
#53	DL#22 [4E-4], DL#16, DL#20, DL#25, DL#6, DL#23, DL#12, DL#18, DL#10, DL#24, DL#26 [4E-2]
#88	DL#20 [4E-5], DL#16, DL#19, DL#26, DL#33 [3E-2]
#94	DL#16 [2E-4], DL#20, DL#26, DL#23, DL#25, DL#5, DL#37, DL#22, DL#19, DL#12, DL#10 [4E-2]

TABLE III: BVs and correlated spatial maps with correlation p-value lower than 0.05. The ranges of the p-values are indicated for each BV.

maps are mostly from discriminative maps $\widetilde{\mathbf{Z}}$ (underlined in Table III), which verifies again that the obtained discriminative maps $\widetilde{\mathbf{Z}}$ are indeed extracting discriminative features, and the common maps $\overline{\mathbf{Z}}$ are indeed obtaining shared features. In fact, most of the BVs are showing significant correlations with DL maps DL#16 (brainstem), DL#22 (sensory motor and motor), DL#37 (anterior cingulate cortex) and DL#26 (associated with logical condition and item recognition). Since the listed BVs are mainly associated with working and visual memory (BV#9, BV#32, BV#53), and verbal and working memory (BV#88, BV#94), finding highly correlated discriminative neural activations in sensory, anterior cingulate cortex, logical condition and item recognition related areas is justified and is in line with many previous studies [63]–[67]. For instance, DL#22, which is showing activations in motor and sensory motor regions that are related to planning, monitoring, decision making and execution of motor activity, is showing high associations with almost all the BVs. DL#37 is a map for anterior cingulate cortex, known to be associated with attention, decision making, emotion, performance monitoring, and error detection. It is also worth noting that DL#25, which is identified by the DL method (see Fig. 8) is highly correlated with BVs #9, #32, #53, and #94, further illustrating the advantage of the proposed discriminative DL method.

V. Conclusion

A novel supervised DL method has been proposed for multi-subject fMRI data analysis to extract brain activation maps that are common across different groups of subjects as well as the maps that are discriminative for predicting group labels. The dictionary and the corresponding sparse matrix have been structured with common and individual submatrices for this purpose, and the labels were incorporated using Fisher's discriminant criterion. Given that spatial functional activations are typically sparse and localized, the sparsity constraint has been imposed on the spatial map factor, and the corresponding weights of different subjects' contributions were used for classification. An optimization algorithm to solve the proposed formulation was derived using alternating minimization based on convex subproblems. The stability of the resulting algorithms was compared to that of the unsupervised DL method, and it was found that our algorithm yielded more stable maps when the same number of components were extracted. The estimated brain maps were also compared carefully with the maps from the ICA. It was observed that the DL formulation not only reproduced most of the ICA components but also extracted components that are more discriminative, including some novel maps that were not discovered by the ICA approach. A correlation analysis with separate behavioral test scores available for the same set of subjects further verified the validity and usefulness of the DL analysis.

Acknowledgments

The authors would like to thank Dr. Vince Calhoun for providing the feature dataset, and Qunfang Long and Suchita Bhinge for their helpful feedback on the analysis results.

References

[1] G. D. Jackson, A. Connelly, J. H. Cross, I. Gordon, and D. G. Gadian, "Functional magnetic resonance imaging of focal seizures," *Neurology*, vol. 44, no. 5, pp. 850–850, 1994.

[2] J. H. Callicott, M. F. Egan, V. S. Mattay, A. Bertolino, A. D. Bone, B. Verchinksi, and D. R. Weinberger, "Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia," *Am. J. Psychiatry*, vol. 160, no. 4, pp. 709–719, 2003.

[3] B. Luna and J. A. Sweeney, "The emergence of collaborative brain function: FMRI studies of the development of response inhibition," *Ann. New York Acad. Sci.*, vol. 1021, no. 1, pp. 296–309, 2004.

[4] J. Taghia, W. Cai, S. Ryali, J. Kochalka, J. Nicholas, T. Chen, and V. Menon, "Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition," *Nat. Comm.*, vol. 9, no. 1, pp. 2505–2523, 2018.

[5] A. A. Phillips, F. H. Chan, M. M. Z. Zheng, A. V. Krassioukov, and P. N. Ainslie, "Neurovascular coupling in humans: Physiology, methodological advances and clinical implications," *J. Cerebral Blood Flow & Metabolism*, vol. 36, no. 4, pp. 647–664, 2016.

[6] R. J. Huster, S. Debener, T. Eichele, and C. S. Herrmann, "Methods for simultaneous EEG-fMRI: An introductory review," *J. Neurosci.*, vol. 32, no. 18, pp. 6053–6060, 2012.

[7] E. Acar, C. Schenker, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data," *Front. Neurosci.*, vol. 13, pp. 416–431, 2019.

[8] J. Xu, M. N. Potenza, and V. D. Calhoun, "Spatial ICA reveals functional activity hidden from traditional fMRI GLM-based analyses," *Front. Neurosci.*, vol. 7, pp. 154–157, 2013.

[9] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional MRI data: An overview," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation*, 2003, pp. 281–288.

[10] S. M. Smith, "Overview of fMRI analysis," *British J. Radiology*, vol. 77, no. suppl_2, pp. S167–S175, 2004.

[11] 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.

[12] 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," *Hum. Brain Mapp.*, vol. 14, no. 3, pp. 140–151, 2001.

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

[14] T. Kim, I. Lee, and T. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, May 2006, pp. 1393–1396.

[15] 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," *Front. Syst. Neurosci.*, vol. 8, pp. 106–123, 2014.

[16] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun, "Deep learning for neuroimaging: A validation study," *Front. Neurosci.*, vol. 8, pp. 229–239, 2014.

[17] S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Proc. Future Tech. Conf.*, San Francisco, CA, Dec. 2016, pp. 816–820.

[18] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, "Brain biomarker interpretation in ASD using deep learning and fMRI," in *Proc. Int. Conf. Med. Imag. Comput. Computer-Assisted Intervention*, Granada, Spain, Sep. 2018, pp. 206–214.

[19] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii, "Deep learning of fMRI big data: A novel approach to subject-transfer decoding," *arXiv preprint arXiv:1502.00093*, 2015.

[20] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neurocomputing*, vol. 15, no. 2, pp. 349–396, 2003.

[21] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neurocomputing*, vol. 12, no. 2, pp. 337–365, 2000.

[22] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[23] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. Conf. Comput. Vis. Pattern Recogn.*, Providence, RI, Jun. 2011, pp. 457–464.

[24] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, 2013.

[25] H. Hu, B. Wohlberg, and R. Chartrand, "Task-driven dictionary learning for inpainting," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 3543–3547.

[26] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. Conf. Comput. Vis. Pattern Recogn.*, San Francisco, CA, Jun. 2010, pp. 2691–2698.

[27] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[28] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1076–1089, May 2011.

[29] J. Lv, X. Jiang, X. Li, D. Zhu, H. Chen, T. Zhang, S. Zhang, X. Hu, J. Han, H. Huang, et al., "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Med. Imag. Anal.*, vol. 20, no. 1, pp. 112–134, 2015.

[30] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Proc. Biennial Int. Conf. Inf. Process. Med. Imag.*, Pacific Grove, CA, Jun.–Jul. 2011, pp. 562–573.

[31] Q. Long, S. Bhinge, Y. Levin-Schwartz, Z. Boukouvalas, V. D. Calhoun, and T. Adalı, "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.

[32] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[33] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Computer Vis.*, Nov. 2011, pp. 543–550.

[34] A. Iqbal, A.-K. Seghouane, and T. Adali, "Shared and subject-specific dictionary learning (ShSSDL) algorithm for multisubject fMRI data analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2519–2528, Nov. 2018.

[35] S. Ghosal, Q. Chen, A. L. Goldman, W. Ulrich, K. F. Berman, D. R. Weinberger, V. S. Mattay, and A. Venkataraman, "A generative-predictive framework to capture altered brain activity in fMRI and its association with genetic risk: Application to schizophrenia," in *Proc. SPIE Med. Imag.: Image Process.*, San Diego, CA, Jun. 2019, vol. 10949, pp. 1–11.

[36] C. F. Beckmann, M. Jenkinson, and S. M. Smith, "General multilevel linear modeling for group analysis in FMRI," *Neuroimage*, vol. 20, no. 2, pp. 1052–1063, 2003.

[37] V. D. Calhoun and T. Adalı, "Feature-based fusion of medical imaging data," *IEEE Trans. Info. Technology in Biomedicine*, vol. 13, no. 5, pp. 711–720, 2008.

[38] V. D. Calhoun, T. Adalı, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson, "A method for multitask fMRI data fusion applied to schizophrenia," *Hum. Brain Mapp.*, vol. 27, no. 7, pp. 598–610, 2006.

[39] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, 2017.

[40] K. Dontaraju, S.-J. Kim, M. Akhonda, and T. Adali, "Capturing common and individual components in fMRI data by discriminative dictionary learning," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2018, pp. 1351–1356.

[41] M. J. McKeown, T.-P. Jung, S. Makeig, G. Brown, S. K. Kindermann, T.-W. Lee, and T. J. Sejnowski, "Spatially independent activity patterns in functional MRI data during the Stroop color-naming task," *Proc. Natl. Acad. Sci.*, vol. 95, pp. 803–810, Feb. 1998.

[42] V. D. Calhoun, P. K. Maciejewski, G. D. Pearlson, and K. A. Kiehl, "Temporal lobe and default hemodynamic brain modes discriminate between schizophrenia and bipolar disorder," *Hum. Brain Mapp.*, vol. 29, no. 11, pp. 1265–1275, 2008.

[43] I. Tosic and P. Frossard, "Dictionary learning: What is the right representation for my signal?," *IEEE Signal Process. Mag.*, vol. 28, pp. 27–38, 2011.

[44] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learning Res.*, vol. 11, no. Jan, pp. 19–60, 2010.

[45] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[46] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.

[47] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Trans. Neural Net. Learn. Syst.*, vol. 27, no. 11, pp. 2160–2173, 2015.

[48] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[49] V. P. Clark, S. Fannon, S. Lai, R. Benson, and L. Bauer, "Responses to rare visual target and distractor stimuli using event-related fMRI," *J. Neurophysiology*, vol. 83, no. 5, pp. 3133–3139, 2000.

[50] R. L. Gollub, J. M. Shoemaker, M. D. King, T. White, S. Ehrlich, S. R. Sponheim, V. P. Clark, J. A. Turner, B. A. Mueller, V. Magnotta, et al., "The MCIC collection: A shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinform.*, vol. 11, no. 3, pp. 367–388, 2013.

[51] A. M. Michael, S. A. Baum, J. F. Fries, B. Ho, R. K. Pierson, N. C. Andreasen, and V. D. Calhoun, "A method to fuse fMRI tasks through spatial correlations: Applied to schizophrenia," *Hum. Brain Mapp.*, vol. 30, no. 8, pp. 2512–2529, 2009.

[52] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia," *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.

[53] W. Du, Y. Levin-Schwartz, G. Fu, S. Ma, V. D. Calhoun, and T. Adalı, "The role of diversity in complex ICA algorithms for fMRI analysis," *J. Neurosci. Methods*, vol. 264, pp. 129–135, 2016.

[54] T. C. Koopmans and M. Beckmann, "Assignment problems and the location of economic activities," *Econometrica*, pp. 53–76, 1957.

[55] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Log. Qtly.*, vol. 2, no. 1-2, pp. 83–97, 1955.

[56] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proc. Amer. Math. Soc.*, vol. 7, no. 1, pp. 48–50, 1956.

[57] X. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, 2010.

[58] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[59] M. Maneshi, S. Vahdat, J. Gotman, and C. Grova, "Validation of shared and specific independent component analysis (SSICA) for between-group comparisons in fMRI," *Front. Neurosci.*, vol. 10, pp. 417–435, 2016.

[60] G. Fu, M. Anderson, and T. Adalı, "Likelihood estimators for dependent samples and their application to order detection," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4237–4244, 2014.

[61] J. Brooks, O. Faull, K. Pattinson, and M. Jenkinson, "Physiological noise in brainstem fMRI," *Front. Human Neurosci.*, vol. 7, pp. 1–13, 2013.

[62] Y. Kopsinis, H. Georgiou, and S. Theodoridis, "FMRI unmixing via properly adjusted dictionary learning," in *Proc. Eur. Signal Process. Conf.*, Lisbon, Portugal, Sep. 2014, pp. 2075–2079.

[63] J. Vannest, J. P. Szaflarski, M. D. Privitera, B. K. Schefft, and S. K. Holland, "Medial temporal fMRI activation reflects memory lateralization and memory performance in patients with epilepsy," *Epilepsy & Behavior*, vol. 12, no. 3, pp. 410–418, 2008.

[64] S. J. Astley, E. H. Aylward, H. C. Olson, K. Kerns, A. Brooks, T. E. Coggins, J. Davies, S. Dorn, B. Gendler, T. Jirikowic, et al., "Functional magnetic resonance imaging outcomes from a comprehensive magnetic resonance study of children with fetal alcohol spectrum disorders," *J. Neurodevelopmental Disorders*, vol. 1, no. 1, pp. 61–80, 2009.

[65] B. D. Bell, "WMS-III logical memory performance after a two-week delay in temporal lobe epilepsy and control groups," *J. Clinical Experimental Neuropsychology*, vol. 28, no. 8, pp. 1435–1443, 2006.

[66] J. L. Cuzzocreo, M. A. Yassa, G. Verduzco, N. A. Honeycutt, D. J. Scott, and S. S. Bassett, "Effect of handedness on fMRI activation in the medial temporal lobe during an auditory verbal memory task," *Hum. Brain Mapp.*, vol. 30, no. 4, pp. 1271–1278, 2009.

[67] R. Massuda, J. Bücker, L. S. Czepielewski, J. C. Narvaez, M. Pedrini, B. T. Santos, A. S. Teixeira, A. L. Souza, M. P. Vasconcelos-Moreno, M. Vianna-Sulzbach, et al., "Verbal memory impairment in healthy siblings of patients with schizophrenia," *Schizophrenia Res.*, vol. 150, no. 2-3, pp. 580–582, 2013.