ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096473

NEW INTERPRETABLE PATTERNS AND DISCRIMINATIVE FEATURES FROM BRAIN FUNCTIONAL NETWORK CONNECTIVITY USING DICTIONARY LEARNING

F. Ghayem[*]*, H. Yang*[*]*, F. Kantar*[*]*, S.-J. Kim*[*]*, 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

Independent component analysis (ICA) of multi-subject functional magnetic resonance imaging (fMRI) data has proven useful in providing a fully multivariate summary that can be used for multiple purposes. ICA can identify patterns that can discriminate between healthy controls (HC) and patients with various mental disorders such as schizophrenia (Sz). Temporal functional network connectivity (tFNC) obtained from ICA can effectively explain the interactions between brain networks. On the other hand, dictionary learning (DL) enables the discovery of hidden information in data using learnable basis signals through the use of sparsity. In this paper, we present a new method that leverages ICA and DL for the identification of directly interpretable patterns to discriminate between the HC and Sz groups. We use multi-subject resting-state fMRI data from 358 subjects and form subject-specific tFNC feature vectors from ICA results. Then, we learn sparse representations of the tFNCs and introduce a new set of sparse features as well as new interpretable patterns from the learned atoms. Our experimental results show that the new representation not only leads to effective classification between HC and Sz groups using sparse features, but can also identify new interpretable patterns from the learned atoms that can help understand the complexities of mental diseases such as schizophrenia.

***Index Terms*—** ICA, dictionary learning, functional network connectivity, multi-subject data, resting-state fMRI

1. INTRODUCTION

An important goal in neuroscience is the development of methods for identifying interpretable patterns that provide discrimination between healthy controls (HC) and different groups of patients. Functional magnetic resonance imaging (fMRI) has proven useful for the study of healthy brain as well as different brain disorders, including schizophrenia (Sz) [1–3]. However, the high dimensionality structure and the noisy nature of fMRI data cause some challenges. Independent component analysis (ICA) has proven particularly useful for feature selection and statistical analysis of fMRI data [4–6].

ICA is a data-driven approach that decomposes the fMRI data into a set of independent components and their corresponding time courses (TC). Unlike classical model-driven methods, such as the general linear model (GLM) [7] that requires predefined model parameters, ICA decomposes brain activities into functional networks that are maximally independent. By concatenating individual subject data and applying one ICA estimation on the aggregate data, group ICA (GICA) [8] generalizes ICA to multi-subject analysis that provides for group inferences. The brain networks that are identified by ICA can be used for studying the inter-network relationships through temporal functional network connectivity (tFNC): the Pearson correlations between pairs of TCs. tFNC has been shown to be highly informative, for instance, chronic mental disorders such as schizophrenia are characterized by significant abnormalities in brain connections [9, 10].

Deep learning methods are also frequently used to detect discriminating biomarkers, but the findings are not directly interpretable and additional steps such as relevance propagation are typically needed [11–14]. In contrast, matrix decompositions such as dictionary learning offer a sparse linear decomposition of the signals based on a set of interpretable bases called atoms [15]. These techniques have shown to extract new, easily comprehensible patterns from the data, revealing data's hidden information [16–21].

In this paper, we develop a method to extract a set of powerful features from resting state fMRI data by bringing together the advantages of ICA and DL. Given that tFNCs are effective in differentiating between the HC and Sz groups, these new features are derived from the sparse representation of the tFNCs from resting-state fMRI data. Our experimental findings show that compared with the original tFNCs, these new features help improve the classification performance between HC and Sz groups. In addition, the approach identifies novel, interpretable biomarkers that help explain the complexities of brain disorders such as schizophrenia.

In the rest of this paper, Section 2 reviews the ICA and DL methods. The proposed framework is discussed in Section 3. Section 4 presents experimental results. The conclusions and perspectives are discussed in Section 5.

2. BACKGROUND

In this section, we first provide a background on how the functional network connectivity features we make use of are created. Then, we review dictionary learning, a central element of our method.

2.1. Functional network connectivity using group ICA

Consider the mixture model $\mathbf{x}(v) = \mathbf{A}\mathbf{s}(v)$, where $\mathbf{x}(v)$ is the observed mixture of N statistically independent signals (components) $\mathbf{s}(v) = [s_1(v) \dots s_N(v)]^T$ at voxel v mixed via $\mathbf{A}$. The components can be estimated as $\hat{\mathbf{s}}(v) = \mathbf{W}\mathbf{x}(v)$, where $\mathbf{W} \in \mathbb{R}^{N \times N}$ is a demixing matrix, which can be obtained by ICA. We make use of GICA [22], where temporal concatenation of subject datasets is

This work was supported in part by NSF-NCS 1631838, 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).

applied to form group data followed by performing one ICA on the group data. The subject-specific result can be achieved by back-reconstruction, which allows the comparison of spatial maps and time courses across subjects, while also addressing the issue with permutation ambiguity inherent in the ICA [23]. The temporal interactions between brain networks for the k^{th} subject can be presented as tFNC$^{[k]} \in \mathbb{R}^{N\times N}$ through the Pearson correlation between time courses from $\mathbf{A}^{[k]}$, which underwent postprocessing procedure in [24]. Since tFNC$^{[k]}$ is symmetric, for subject k, we only use its upper triangular data, and arrange them in vector $\mathbf{f}^{[k]}$ of size $P = \frac{N(N-1)}{2}$ to serve as the subject's *FNC-feature* vector.

2.2. Dictionary learning and sparse representation

Dictionary learning is the task of estimating a set of basis signals, called atoms, using a training dataset such that each training sample can be written as a sparse linear combination of the learned atoms. Mathematically, consider some training feature vectors $\mathbf{f}^{[k]} = [f_1^{[k]}, \ldots, f_P^{[k]}]^T$ collected as the columns of the matrix $\mathbf{F} = [\mathbf{f}^{[1]}, \ldots, \mathbf{f}^{[K]}] \in \mathbb{R}^{P\times K}$. Then, the goal is to learn a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_G] \in \mathbb{R}^{P\times G}$, with G as the number of atoms, such that $\mathbf{F} = \mathbf{DZ}$ and the columns of the coefficient matrix $\mathbf{Z} = [\mathbf{z}^{[1]}, \ldots, \mathbf{z}^{[K]}] \in \mathbb{R}^{G\times K}$ are sparse. That is, each feature vector $\mathbf{f}^{[k]}$ can be written as $\mathbf{f}^{[k]} = \sum_{i=1}^{G} z_g \mathbf{d}_g$, where most of z_g's are zeros. To learn $\mathbf{D}$, a sparsity promoting function denoted r is considered to impose a sparsity constraint on the coefficient matrix $\mathbf{Z}$. The dictionry learning problem is then formulated as follows [15]:

$$\min_{\mathbf{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{F} - \mathbf{DZ}\|_F^2 + \lambda \cdot r(\mathbf{Z}),$$
$$\text{s.t. } \mathbf{D} \in \mathcal{D} := \{\mathbf{D} : \|\mathbf{d}_g\|_2 = 1, g = 1, 2, \ldots, G\}, \qquad \text{(P1)}$$

where $\|.\|_F$ is the Frobenius norm, and $\lambda > 0$ is a sparsity parameter.

Problem (P1) can be solved using alternating minimization, which iterates between two steps: dictionary update (DU), and sparse representation (SR). The DU step minimizes (P1) across dictionary $\mathbf{D}$ while assuming that $\mathbf{Z}$ is fixed, and in (SR) step, (P1) is solved with respect to $\mathbf{Z}$ assuming $\mathbf{D}$ is fixed. Different approaches can be used to alternate between these two steps [15, 25–27].

3. METHODOLOGY

In this section, we propose the sparse representation of the FNC-features from fMRI data, and show how it reveals new interpretable patterns that can provide discrimination between HC and Sz.

3.1. Joint classifier and dictionary learning for FNC-features

We propose to jointly learn a linear classifier and a dictionary to find the sparse representations of the FNC-feature vectors $\mathbf{f}^{[k]} = \mathbf{Dz}^{[k]}$. We concatenate the feature vectors of all subjects in matrix $\mathbf{F} = [\mathbf{f}^{[1]}, \ldots \mathbf{f}^{[K]}]$, and consider a common dictionary $\mathbf{D} \in \mathbb{R}^{P\times G}$ with G as the number of atoms. Then, the DL problem will be as in (P1). To jointly learn a linear classifier with the dictionary $\mathbf{D}$, we use the binary labels $\mathbf{l}^{(\text{HC})} = [0, 1]^T$ and $\mathbf{l}^{(\text{Sz})} = [1, 0]^T$ for HC and Sz groups, respectively. So, with a linear classifier $\mathbf{W} \in \mathbb{R}^{2\times G}$, the label of subject k can be estimated as $\mathbf{l}^{[k]} = \mathbf{Wz}^{[k]}$. We denote the data corresponding to training and test sets with subscript "tr", and "ts", respectively. By concatenating the labels for K_{tr} training datasets in the label matrix $\mathbf{L}_{\text{tr}} = [\mathbf{l}_{\text{tr}}^{[1]}, \ldots, \mathbf{l}_{\text{tr}}^{[K_{\text{tr}}]}]$, we have $\mathbf{L}_{\text{tr}} = \mathbf{WZ}_{\text{tr}}$. Inspired by [28], we add this equality as a discriminative penalty with parameter β to the cost function in (P1), and solve the following problem:

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \frac{1}{2}\|\mathbf{F} - \mathbf{DZ}\|_F^2 + \lambda \cdot r(\mathbf{Z}) + \frac{\beta}{2}\|\mathbf{L}_{\text{tr}} - \mathbf{WZ}_{\text{tr}}\|_F^2,$$
$$\text{s.t. } \mathbf{D} \in \mathcal{D} := \{\mathbf{D} : \|\mathbf{d}_g\|_2 = 1, g = 1, 2, \ldots, G\}. \qquad \text{(P2)}$$

We note that in the above formulation, we consider to update the dictionary $\mathbf{D}$ using the whole dataset $\mathbf{F} = [\mathbf{F}_{\text{tr}}, \mathbf{F}_{\text{ts}}]$ and $\mathbf{Z} = [\mathbf{Z}_{\text{tr}}, \mathbf{Z}_{\text{ts}}]$, while for the classification term, we only involve the training labels.

To solve (P2), we perform alternating minimization. At iteration i, dictionary update (DU) is obtained by setting the gradient of the target function with respect to $\mathbf{D}$ to zero, and projecting the result to the set $\mathcal{D}$. This results in a closed form expression (DU): $\mathbf{D}^{(i)} \leftarrow \mathcal{P}_{\mathcal{D}}\Big(\mathbf{FZ}^T(\mathbf{ZZ}^T)^{-1}\Big)$, where $\mathcal{P}_{\mathcal{D}}$ is the projection on the set $\mathcal{D}$.

In the sparse representation (SR) step, different approaches such as orthogonal matching pursuit (OMP) and proximal methods can be used [15, 25]. Here, we use iterative proximal-projection approach due to its flexibility to a range of sparsity-promoting functions, including non-convex and non-smooth scenarios [25]. The *proximal mapping* is a key operator in these algorithms defined as:

Definition 1. [29] *The proximal mapping of a proper and lower semicontinuous function* $r : \text{dom}_r \to (-\infty, +\infty]$ *at* $\mathbf{x} \in \mathbb{R}^n$ *is* $\text{prox}_r(\mathbf{x}) = \text{argmin}_{\mathbf{u}\in\text{dom}_r} \{\frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_F^2 + r(\mathbf{u})\}$.

We first decompose the sparse coefficients into training and testing sets, and we separately update $\mathbf{Z}_{\text{tr}}$ and $\mathbf{Z}_{\text{ts}}$ using proximal method. The proximal approach consists of two steps: 1) gradient step, 2) sparsification step. We start with updating $\mathbf{Z}_{\text{ts}}$. By defining $\mathbf{J}_{\mathbf{Z}_{\text{ts}}} \triangleq \frac{1}{2}\|\mathbf{F}_{\text{ts}} - \mathbf{DZ}_{\text{ts}}\|_F^2$, the gradient step is $\nabla\mathbf{J}_{\mathbf{Z}_{\text{ts}}} = -\mathbf{D}^T\mathbf{F}_{\text{ts}} + \mathbf{D}^T\mathbf{DZ}_{\text{ts}}$. Then, the sparsification step is (SR-sparsification): $\mathbf{Z}_{\text{ts}} \leftarrow \text{prox}_{\mu\lambda r(\cdot)}(\mathbf{Z}_{\text{ts}} - \mu\nabla\mathbf{J}_{\mathbf{Z}_{\text{ts}}})$, where $\text{prox}_g(\cdot)$ is given in Definition 1. Selecting various functions for $r(\cdot)$ results in different sparsifications. For example, the SR-sparsification step will be soft-thresholding if we use the ℓ_1-norm, and hard-thresholding if we use the ℓ_0-norm. In this paper, we consider a κ-sparse constraint, and set

$$r(\mathbf{Z}_{\text{ts}}) \triangleq \begin{cases} 0 & \text{if } \|\mathbf{Z}_{\text{ts}}\|_0 \le \kappa \\ \infty & \text{o.w.} \end{cases}. \qquad (1)$$

Algorithm 1 Proposed method for solving (P2)

Inputs: $\mathbf{F}_{\text{tr}}$, $\mathbf{F}_{\text{ts}}$, $\mathbf{L}_{\text{tr}}$, κ, G, μ, β, Iter$_{\text{in}}$, Iter$_{\text{out}}$
Initialization: $\mathbf{D}$ & $\mathbf{W}$: DCT , $\mathbf{Z} = [0]_{G\times K}$
for $i = 1 :$ Iter$_{\text{out}}$ **do**
  for $j = 1 :$ Iter$_{\text{in}}$ **do**
    SR-gradient:
      $\nabla\mathbf{J}_{\mathbf{Z}_{\text{ts}}} = -\mathbf{D}^T\mathbf{F}_{\text{ts}} + \mathbf{D}^T\mathbf{DZ}_{\text{ts}}$
      $\nabla\mathbf{J}_{\mathbf{Z}_{\text{tr}}} = -\mathbf{D}^T(\mathbf{F}_{\text{tr}} - \mathbf{DZ}_{\text{tr}}) - \beta\mathbf{W}^T(\mathbf{L}_{\text{tr}} - \mathbf{WZ}_{\text{tr}})$
    SR-sparsification:
      $\mathbf{Z}_{\text{ts}} \leftarrow \text{prox}_{\mu\lambda r(\cdot)}(\mathbf{Z}_{\text{ts}} - \mu\nabla\mathbf{J}_{\mathbf{Z}_{\text{ts}}})$
      $\mathbf{Z}_{\text{tr}} \leftarrow \text{prox}_{\mu\lambda r(\cdot)}(\mathbf{Z}_{\text{tr}} - \mu\nabla\mathbf{J}_{\mathbf{Z}_{\text{tr}}})$
  end for
  $\mathbf{Z} \leftarrow [\mathbf{Z}_{\text{tr}}, \mathbf{Z}_{\text{ts}}]$
  DU: $\mathbf{D} \leftarrow \mathcal{P}_{\mathcal{D}}\Big(\mathbf{FZ}^T(\mathbf{ZZ}^T)^{-1}\Big)$
  Classifier update: $\mathbf{W} \leftarrow \mathbf{L}_{\text{tr}}\mathbf{Z}_{\text{tr}}^T(\mathbf{Z}_{\text{tr}}\mathbf{Z}_{\text{tr}}^T)^{-1}$
end for
Output: $\mathbf{D}$, $\mathbf{Z}$, $\mathbf{W}$

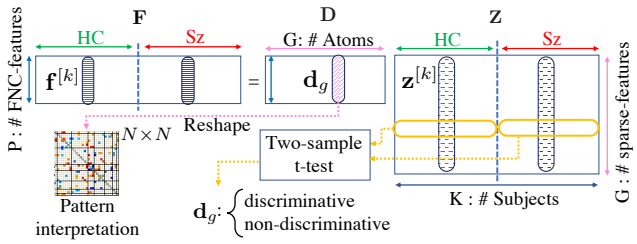


Fig. 1: Proposed framework.

The proximal of the above function is a projection operation that keeps the κ largest elements of $|\mathbf{Z}_{\text{ts}}|$ with setting the rest to 0.

Now, we define $\mathbf{J}_{\mathbf{Z}_{\text{tr}}} \triangleq \frac{1}{2}\|\mathbf{F}_{\text{tr}} - \mathbf{D}\mathbf{Z}_{\text{tr}}\|_F^2 + \frac{\beta}{2}\|\mathbf{L}_{\text{tr}} - \mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2$ for the update of $\mathbf{Z}_{\text{tr}}$. With a similar procedure, we obtain the update as (SR-gradient): $\nabla \mathbf{J}_{\mathbf{Z}_{\text{tr}}} = -\mathbf{D}^T(\mathbf{F}_{\text{tr}} - \mathbf{D}\mathbf{Z}_{\text{tr}}) - \beta \mathbf{W}^T(\mathbf{L}_{\text{tr}} - \mathbf{W}\mathbf{Z}_{\text{tr}})$, and (SR-sparsification): $\mathbf{Z}_{\text{tr}} \leftarrow \text{prox}_{\mu\lambda r(\cdot)}(\mathbf{Z}_{\text{tr}} - \mu \nabla \mathbf{J}_{\mathbf{Z}_{\text{tr}}})$.

The update of the linear classifier $\mathbf{W}$ is obtained by setting the gradient of the target function in (P2) with respect to $\mathbf{W}$ to zero. This results in the closed form expression: $\mathbf{W} \leftarrow \mathbf{L}_{\text{tr}}\mathbf{Z}_{\text{tr}}^T(\mathbf{Z}_{\text{tr}}\mathbf{Z}_{\text{tr}}^T)^{-1}$. The final algorithm is summarized in Algorithm 1.

3.2. Interpretability, and discrimination in sparse space

In this part, from the sparse decomposition of the tFNCs, we introduce new discriminative features, and new interpretable patterns. To accomplish this goal, we suggest assigning the new *sparse-feature* vector of subject k as $\mathbf{z}^{[k]} = [z_1^{[k]}, \ldots, z_G^{[k]}]^T$, corresponding to the k^{th} columns of $\mathbf{Z}$. Each element $z_g^{[k]}$ describes the contribution of the g^{th} atom $\mathbf{d}_g$ to the representation of the initial tFNC-feature vector $\mathbf{f}^{[k]}$. On the other hand, every atom $\mathbf{d}_g$ has the same dimensionality as the original tFNC-feature vectors. Therefore, by rearranging the atoms into a symmetric matrix of size $N \times N$, we can interpret the resulting matrices as in the tFNCs. This interpretation again explains the interaction between brain networks $\{\mathbf{s}_1^{[k]}, \ldots, \mathbf{s}_N^{[k]}\}$ where $\mathbf{s}_n^{[k]} = [s_n^{[k]}(v)]^T$ for $v = 1, \ldots, V$ representing the voxels, but this time through the new patterns obtained from the atoms.

Now, we can perform a statistical analysis to determine which of these atoms provide discrimination between the two groups. For example, if we divide the columns of the sparse coefficient matrix into the HC and Sz groups as $\mathbf{Z} = [\mathbf{Z}^{[\text{HC}]}, \mathbf{Z}^{[\text{Sz}]}]$, we can determine if the pattern corresponding to the g^{th} atom is discriminating or not by performing a two-sample t-test [30] between the g^{th} rows of $\mathbf{Z}^{[\text{HC}]}$ and $\mathbf{Z}^{[\text{Sz}]}$. Fig. 1 illustrates the steps for interpretation as well as a comparison of the FNC-feature with the sparse-feature vector.

4. EXPERIMENTAL RESULTS

4.1. Data preparation

Extraction of FNC-features. We use multi-subject resting state fMRI (rs-fRMI) data from the bipolar and schizophrenia network for intermediate phenotypes (BSNIP) dataset [31] considering 179 healthy controls (HC) and 179 patients with schizophrenia (Sz), using five sites: Baltimore, Chicago, Dallas, Detroit, and Hartford. All images were collected from a single 5-min run on a 3-T scanner and all subjects were instructed to have their eyes open and remain still during the entire scan. The fMRI data were then resampled to $3 \times 3 \times 3$ mm^3 isotropic voxels and smoothed using a Gaussian kernel with a full width at half maximum (FWHM) = 6 mm. Only the subjects who passed quality control [32] were selected. We removed the first three timepoints for the following ICA analysis. Group ICA-EBM [33] is performed to obtain the subject-specific tFNC-feature vectors $\mathbf{f}^{[k]}$. The order is determined as $N = 55$ using the method proposed in [34]. Compared with other ICA algorithms, ICA-EBM has the flexibility of estimating sources from different distributions by using a few classes of nonlinear functions. Out of the 55 estimated components, we selected $N = 32$ as functionally relevant. The size of the tFNC-features was then calculated as $P = \frac{N(N-1)}{2} = 496$.[1]

Extraction of sparse-features. We obtain subject-specific sparse-feature vectors $\mathbf{z}^{[k]}$ by applying the DL approach presented in Section 2.2 on the tFNC-features $\mathbf{f}^{[k]}$. We initialize $\mathbf{D}$ and $\mathbf{W}$ with DCT dictionary [15], and $\mathbf{Z}$ with a null (zero) matrix. We consider a complete dictionary of size $G = P = 496$, and the sparsity level is set to κ-sparse = 50%. The gradient descent step size is $\mu = 0.005$, and the number of inner iteration and outer iteration are set to $\text{Iter}_{\text{in}} = 5$, and $\text{Iter}_{\text{out}} = 200$, respectively. These values are selected empirically based on our observation regarding the convergence behaviour of the sparse representation of the signals. We consider two different scenarios: 1) $\beta = 0$ which represents DL without learning a linear classifier, and 2) $\beta = 0.05$ which considers the linear classifier to be jointly learned with the dictionary. We randomly select 20% of the subjects within each group as test set $\mathbf{F}_{\text{ts}}$, and we keep the rest of the data (80%) as the training set $\mathbf{F}_{\text{tr}}$. With the above setup, we run Algorithm 1 and obtain the sparse coefficients for the subjects in groups HC and Sz, *i.e.* $\mathbf{Z}^{[\text{HC}]}$ and $\mathbf{Z}^{[\text{Sz}]}$.

Table 1: Average classification rates [%].

Metric\Feature	tFNC	Sparse ($\beta = 0$)	Sparse ($\beta = 0.05$)
Recall	74.75 ± 0.61	73.56 ± 0.65	$\mathbf{75.19} \pm 0.65$
Specificity	73.78 ± 0.70	74.14 ± 0.70	$\mathbf{74.47} \pm 0.68$
Precision	74.35 ± 0.50	74.27 ± 0.53	$\mathbf{74.93} \pm 0.51$
Accuracy	74.26 ± 0.40	73.85 ± 0.45	$\mathbf{74.83} \pm 0.43$
F1-score	74.35 ± 0.40	73.72 ± 0.46	$\mathbf{74.87} \pm 0.45$

4.2. Classification results

In this section, we compare the performance of the tFNC-features and sparse-features in the classification of HC and Sz groups. In order to achieve this, we train SVM classifiers [35] with polynomial kernels of order 3, which according to our experiments, provided the best overall performance. In the training phase, we separately use features ($\mathbf{F}_{\text{tr}}$,$\mathbf{L}_{\text{tr}}$) and ($\mathbf{Z}_{\text{tr}}$,$\mathbf{L}_{\text{tr}}$) to train SVM classifiers SVM^{FNC} and SVM^{SPR} using sparse-features and FNC-features, respectively. Then, in the test phase, the test sets $\mathbf{F}_{\text{ts}}$ and $\mathbf{Z}_{\text{ts}}$ are respectively given to SVM^{FNC} and SVM^{SPR}, in order to estimate the labels of the test sets $\widehat{\mathbf{L}}_{\text{ts}}^{\text{FNC}}$ and $\widehat{\mathbf{L}}_{\text{ts}}^{\text{SPR}}$. Comparing the estimated test labels with the actual test label matrix $\mathbf{L}_{\text{ts}}$, we evaluate the classification performance using 5 metrics: recall, specificity, precision, accuracy, and F1-score. By repeating the classification experiment 100 times with new random samples from the training and test sets, we report the average classification rates in Table 1. We see that all classification metrics are improved by using sparse features derived from the dictionary that are jointly learnt with the linear classifier. In addition, as we address below, the decomposition provides better interpretability *w.r.t.*

[1]The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS-0821258, CNS-1228778, and OAC-1726023) and the SCREMS program (grant no. DMS-0821311).

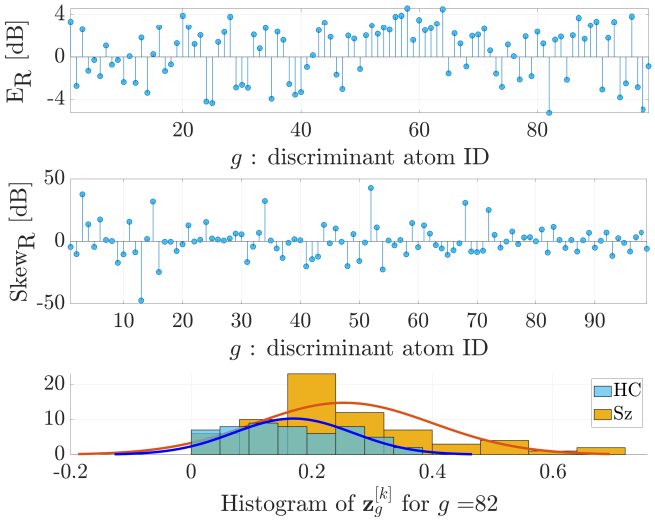
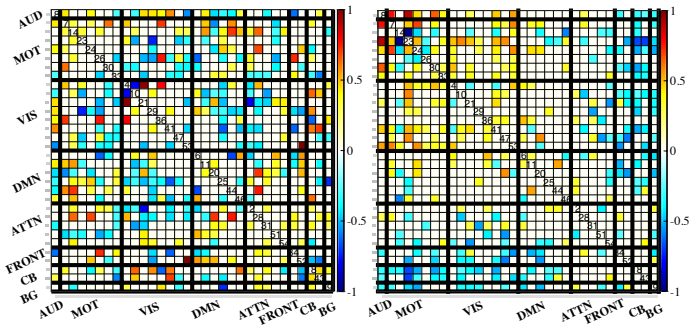


Fig. 2: Statistical analysis of the sparse coefficient corresponding to the 99 discriminative atoms obtained from the two-sample t-test between the sparse coefficients $\mathbf{z}_g^{[\text{HC}]}$ and $\mathbf{z}_g^{[\text{Sz}]}$. Top: the ratio between the average energy and, middle: the ratio between the skewness of $\mathbf{z}_g^{[\text{HC}]}$ and $\mathbf{z}_g^{[\text{Sz}]}$. Bottom: the histogram of $\mathbf{z}_g^{[\text{HC}]}$ and $\mathbf{z}_g^{[\text{Sz}]}$ for $g = 82$.



(a) $\sum \mathbf{d}_g \bar{\mathbf{z}}_g^{[\text{Sz}]}$ for g's dominant in Sz (b) $\sum \mathbf{d}_g \bar{\mathbf{z}}_g^{[\text{Sz}]}$ for g's dominant in HC

Fig. 3: Weighted average of the discriminative atoms that are a) dominant in HC, and b) dominant in Sz groups. More modularity is observed in HC.

Sz and HC differences.

4.3. Discriminant atoms and their interpretability

In order to find the discriminative atoms and tFNC-features between groups HC and Sz, we apply two-sample t-test followed by false discovery rate (FDR) correction [36]. For the FNC-features, the two-sample t-test is applied on the p^{th} feature vectors $\mathbf{f}_p^{[\text{HC}]}$ and $\mathbf{f}_p^{[\text{Sz}]}$, which correspond to the subject indices in groups HC and Sz, respectively. We found 113 discriminative FNC-features out of $P = 496$. We repeated the two-sample t-test for the sparse coefficients corresponding to the g^{th} atom $\mathbf{z}_g^{[\text{HC}]}$ and $\mathbf{z}_g^{[\text{Sz}]}$ (see Fig. 1), and we identified 99 atoms (patterns) that discriminate between HC and Sz groups.

The sparse coefficients provide us with a statistical population that can be further analyzed for better understanding of the contribution of each discriminative atoms in the two groups. In Fig. 2, the top plot shows the ratio between the average energy of the sparse coefficients for HC and Sz groups *i.e.* $\text{E}_\text{R} = 10\log(||\mathbf{z}_g^{[k_\text{HC}]}||_2^2/||\mathbf{z}_g^{[k_\text{Sz}]}||_2^2)$. Here, g represents one of the 99 discriminant atoms specified by the two-sample t-test. The atoms with energy ratios above 0 are those with higher energies for the HC group, which indicates that they dominant in the HC group, *i.e.*, contribute more to the representation of tFNC-features for the HC group. The discriminative atoms that are dominant in Sz group are also indicated by points below the threshold of zero. Also, a higher absolute value $|\text{E}_\text{R}|$ indicates a higher energy difference between the two groups. The middle plot in Fig. 2 reports the ratio between the skewness of the sparse coefficients for HC and Sz groups, *i.e.*, $\text{Skew}_\text{R} = 10\log(|\text{skew}(\mathbf{z}_g^{[k_\text{HC}]})|/|\text{skew}(\mathbf{z}_g^{[k_\text{Sz}]})|)$. Similarly, atoms with positive values of Skew_R are more skewed in the distribution of $\mathbf{z}_g^{[\text{HC}]}$, and vice versa for the atoms with negative Skew_R. A larger absolute value $|\text{Skew}_\text{R}|$ indicates more skewness difference between the two groups. The bottom plot in Fig. 2 shows the histogram of the sparse coefficients for a sample discriminative atom $g = 64$. From the histogram, we can visually verify that the sparse coefficients corresponding to this atom have larger energy in Sz group and the skewness in the distribution of the sparse coefficients is higher in this group.

Besides information from the statistical analysis of the sparse coefficients, we can interpret the discriminative atoms by reshaping them to a symmetric matrix of size $N \times N$ which has the same shape as the tFNC. Fig. 3 shows the average pattern corresponding to the first 15 dominant atoms in Sz and HC groups with the largest energy ratios (according to Fig. 2-top). The results are shown with ae threshold level of 0.25. The pattern in Fig. 3-(a) constitutes 6.7% of the overall energy in $\mathbf{Z}^{[\text{HC}]}$, while it is only 3.5% in $\mathbf{Z}^{[\text{Sz}]}$. These contributions for Fig. 3-(b), are 1.3% and 2.7% for HC and Sz groups, respectively. We note that there is a large difference in the energy between the two groups. This increases our confidence about the discrimination of these patterns between the two groups. in Fig. 3 we can see that these atoms show different patterns in terms of the interaction between different brain networks. Comparing Fig. 3 (a) and (b), we can see more modularity in HC with a swath of negative values in between sensory and DMN to subcortical and frontal regions, which suggests that SZ appears less anatomically organized and structured with more extreme values, as also observed in the histogram Fig. 2-bottom.

5. CONCLUSIONS AND PERSPECTIVES

In this paper, we presented the sparse representation of the subject-specific brain temporal functional network connectivity obtained from independent component analysis of the resting-state multi-subject fMRI dataset. To this end, we suggested to jointly learn a dictionary for the sparse representation of the tFNC features and a linear classifier to determine whether the subjects should be classified as HC or Sz using sparse coefficients as features. Compared with the FNC features, using sparse features, the classification rates improve. More importantly, we identify new discriminative patterns formed from dictionary atoms that can be interpreted as tFNC features, *i.e.*, revealing patterns of interaction between brain networks. This work also provides new perspectives for studying dynamics of fMRI to further investigate brain functionality. Also, learning a non-linear classifier jointly with the dictionary can be used to further improve the classification rates [37], and the approach can be easily extended to multiple sets of fMRI data [38, 39].

6. REFERENCES

[1] J. M. Levin, M. H. Ross, and P. F. Renshaw, "Clinical applications of functional MRI in neuropsychiatry," *J. Neuropsych. Clin. Neurosci.*, vol. 7, pp. 511–522, 1995.

[2] 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, pp. 1265–1275, 2008.

[3] O. Demirci, V. P. Clark, V. A. Magnotta, N. C. Andreasen, J. Lauriello, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun, "A review of challenges in the use of fMRI for disease classification/characterization and a projection pursuit application from a multi-site fMRI schizophrenia study.," *Brain Imag. Behav.*, vol. 2, pp. 207–226, 2008.

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

[5] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components.," *Hum. Brain Mapp.*, vol. 6, pp. 160–188, 1998.

[6] J. I. Arribas, V. D. Calhoun, and T. Adali, "Automatic bayesian classification of healthy controls, bipolar disorder and schizophrenia using intrinsic connectivity maps from fMRI data.," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2850–2860.

[7] S. M. Smith, "Overview of fMRI analysis," *Brit. J. Radiol*, vol. 77, pp. S167–S175, 2004.

[8] 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.

[9] W. Du et al., "High classification accuracy for schizophrenia with rest and task fMRI data," *Front. Hum. Neurosci.*, vol. 6, 2012.

[10] A. M. Shashwath et al., "Differences in resting-state functional magnetic resonance imaging functional network connectivity between schizophrenia and psychotic bipolar probands and their unaffected first-degree relatives," *Biolog. Psych.*, vol. 71, no. 10, pp. 881–889, 2012.

[11] G. Montavon, W. Samek, and K. Müller, "Methods for interpreting and understanding deep neural networks," *Digital signal processing*, vol. 73, pp. 1–15, 2018.

[12] J. Kim, V. D. Calhoun, E. Shim, and J. H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *NeuroImage*, vol. 124, pp. 127–146, 2016.

[13] S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize alzheimer's disease using fMRI data," *Future Technologies Conference*, pp. 816–820, 2016.

[14] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, "Brain biomarker interpretation in asd using deep learning and fMRI," *Int. Conf. Med. Imag. Comput.*, pp. 206–214, 2018.

[15] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[16] P. Comon and C. Jutten, *Handbook of Blind Source Separation*, Elsevier, 2010.

[17] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[18] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[19] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[20] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Applied and Num. Harmonic Anal. Birkhäuser Basel, 2013.

[21] R. Jin et al., "Dictionary learning-based fMRI data analysis for capturing common and individual neural activation maps," *IEEE Journal of Selected Topics in Sig. Proc.*, vol. 14, no. 6, pp. 1265–1279, 2020.

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

[23] Vince D Calhoun, Jingyu Liu, and Tülay Adalı, "A review of group ica for fmri data and ica for joint inference of imaging, genetic and erp data," *NeuroImage*, vol. 45, no. 1, pp. S163–S172, 2009.

[24] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.

[25] F. Ghayem, M. Sadeghi, M. Babaie-Zadeh, S. Chatterjee, M. Skoglund, and C. Jutten, "Sparse signal recovery using iterative proximal projection," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 879–894, 2018.

[26] F. Ghayem, M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Accelerated dictionary learning for sparse signal representation," *International Conf. on LVA and Signal Separation*, pp. 531–541, 2017.

[27] M. Sadeghi, F. Ghayem, M. Babaie-Zadeh, S. Chatterjee, M. Skoglund, and C. Jutten, "L0soft: ℓ_0 minimization via soft thresholding," *European Signal Processing Conference (EUSIPCO)*, 2019.

[28] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *IEEE Computer Society Conf. on CVPR*. IEEE, 2010, pp. 2691–2698.

[29] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2014.

[30] Student, "The probable error of a mean," *Biometrika*, pp. 1–25, 1908.

[31] C. A. Tamminga et al., "Clinical phenotypes of psychosis in the bipolar-schizophrenia network on intermediate phenotypes (B-SNIP)," *American Journal of psychiatry*, vol. 170, no. 11, pp. 1263–1274, 2013.

[32] Y. Du et al., "Evidence of shared and distinct functional and structural brain signatures in schizophrenia and autism spectrum disorder," *Communications biology*, vol. 4, no. 1, pp. 1–16, 2021.

[33] X. L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5151–5164, 2010.

[34] G. S. Fu, M. Anderson, and T. Adalı, "Likelihood estimators for dependent samples and their application to order detection," *IEEE trans. on signal processing*, vol. 62, no. 16, pp. 4237–4244, 2014.

[35] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[36] Y. Benjamini and D. Yekutieli, "False discovery rate–adjusted multiple confidence intervals for selected parameters," *JASA*, vol. 100, no. 469, pp. 71–81, 2005.

[37] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. Bach, "Supervised dictionary learning," *Advances in neural information processing systems*, vol. 21, 2008.

[38] 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. on Medical Imaging*, vol. 30, no. 5, pp. 1076–1089, 2010.

[39] A. K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Transactions on image Processing*, vol. 26, no. 6, pp. 3002–3015, 2017.