Song, L., Ren, Y., Shuhan, X., Hou, Y. & He, X. (2023). A hybrid spatio-temporal deep belief network and sparse representation based framework reveals multi-level core functional components in decoding multi-task fMRI signals. *Network Neuroscience*, Advance publication. https://doi.org/10.1162/netn_a_00334.

A Hybrid Spatio-Temporal Deep Belief Network and Sparse Representation-Based Framework Reveals Multi-Level Core Functional Components in Decoding Multi-Task fMRI Signals

Limei Song[1#], Yudan Ren[1#], Shuhan Xu[1], Yuqing Hou[1], Xiaowei He[1]*

[1] School of Information Science & Technology, Northwest University, China;

[#] These authors contributed equally to this work and should be considered co-first authors.

[*] Corresponding authors.

Abstract

Decoding human brain activity on various task-based functional brain imaging data is of great significance for uncovering the functioning mechanism of the human mind. Currently, most feature extraction model-based methods for brain state decoding are shallow machine learning models, which may struggle to capture complex and precise spatio-temporal patterns of brain activity from the highly noisy fMRI raw data. Moreover, although decoding models based on deep learning methods benefit from their multi-layer structure that could extract spatio-temporal features at multi-scale, the relatively large populations of fMRI datasets are indispensable and the explainability of their results is elusive. To address the above problems, we proposed a computational framework based on hybrid spatio-temporal deep belief network and sparse representations to differentiate multi-task fMRI (tfMRI) signals. Using a relatively small cohort of tfMRI data as a testbed, our framework can achieve an average classification accuracy of 97.86% and define the multi-level temporal and spatial patterns of multiple cognitive tasks. Intriguingly, our model can characterize the key components for differentiating the multi-task fMRI signals. Overall, the proposed framework can identify the interpretable and discriminative fMRI composition patterns at multiple scales, offering an effective methodology for basic neuroscience and clinical research with relatively small cohorts.

Keywords: Multi-task classification, Task-based fMRI, Deep belief network, Sparse representation, Functional brain network.

Introduction

For years, researchers have been attempting to decode the human brain states based on functional magnetic resonance imaging (fMRI) data (Haynes & Rees, 2006; Jang, Plis, Calhoun, & Lee, 2017; Rubin et al., 2017; Stanislas Dehaene, 1998), where distinguishing different cognitive tasks from fMRI data and extracting discriminative fMRI composition patterns are effective means to improve our understanding of the relationship among current cognitive tasks, brain responses, and individual behavior (Friston, 2009; Logothetis, 2008). To decode meaningful neurological patterns embedded in diverse task-based fMRI data, various computational and statistical methods have been proposed in the last decades. The most widely used brain state decoding strategy is multi-voxel pattern analysis (MVPA) (Davatzikos et al., 2005; Jang et al., 2017; Kriegeskorte & Bandettini, 2007). Despite its popularity, its commonly-used classification strategy support vector machine (SVM) usually struggles to perform well on high-dimensional fMRI data and thus requires effective techniques for feature selection/extraction (LeCun, Bengio, & Hinton, 2015; Vieira, Pinaya, & Mechelli, 2017). Hence, the feasibility of feature selection/extraction has been investigated using various machine learning methods (LeCun et al., 2015; Vieira et al., 2017; S. Zhang et al., 2016).

However, most of these machine learning methods rely on shallow models, and their shallow nature may hinder them from effectively capturing non-linear relationships in the highly noisy fMRI raw data, resulting in difficulties in extracting complex and specific spatio-temporal features (Qiang et al., 2020; Rashid, Singh, & Goyal, 2020; Varoquaux & Thirion, 2014).

Recently, studies applying deep learning models such as deep neural network (DNN) and convolutional neural networks (CNN) to decode brain states based on task-based fMRI signals have been reported (J. Hu et al., 2019; Liu, He, Chen, & Gao, 2019; Sotetsu Koyamadaa, 2015; Y. Zhang, Tetrel, Thirion, & Bellec, 2021). Such deep learning models take the advantage of being a multi-layer architecture by stacking multiple building blocks with similar structure, which has demonstrated the ability to significantly reduce noises in raw fMRI data and model the non-linear relationships among neural activities of brain regions, allowing for the extraction of multi-level spatio-temporal features (Bengio, Courville, & Vincent, 2012; Najafabadi et al., 2015; Ren, Xu, Tao, Song, & He, 2021). Nevertheless, there are still some limitations in current brain state decoding strategies based on deep learning models. First, as large-size samples are indispensable for the deep learning model, current decoding models are not suitable for small datasets (Bo Liu, 2017; Litjens et al., 2017; Wang et al., 2020; Wen et al., 2018). For example, Wang et al. (2020) proposed a DNN-based model for tfMRI signal classification, which requires 1034 subjects, making it less practical for clinical populations. Second, most of the decoding models based on deep learning are end-to-end learning and the explainability of such models is elusive (J. Hu et al., 2019; LeCun et al., 2015; Wang et al., 2020). Recently, some researchers have attempted to define the key components for decoding brain states using the machine learning method. For example, our previous study based on sparse dictionary learning has determined that the key components for multi-task classification tend to be functional brain networks (FBNs) (Song, Ren, Hou, He, & Liu, 2022). Another research has shown that artifact components such as movement-related artifacts are significantly more informative with respect to the classification accuracy of the multi-task electroencephalogram (EEG) signals (McDermott et al., 2021). However, uncovering the interpretable key features in decoding tfMRI signals has received much less attention.

Due to the pitfalls in existing research, it is desirable to develop an appropriate framework capable of identifying the interpretable and discriminative fMRI composition patterns embedded in multi-task fMRI data. Thus, in this study, we aim to extract both multi-level group-wise temporal features and spatial features from tfMRI signals, and define interpretable classification features for multi-task fMRI data simultaneously. Recent studies have revealed that the deep belief network (DBN) can effectively identify multi-layer spatial and temporal features from fMRI signals (Dong, 2020; Ren et al., 2021), which is typically stacked by multiple Boltzmann machine (RBM) (Geoffrey E Hinton & Sejnowski, 1986) and thus can naturally act as a multi-level feature extractor. Furthermore, these prior studies have integrated the least absolute shrinkage and selection operator (LASSO) regression with the DBN model, indicating the efficacy of LASSO regression in extracting relevant spatial patterns. Thus, we here proposed a novel two-stage feature extraction framework based on hybrid DBN and sparse representations framework (DBN-SR) to decode multi-task fMRI signals with the capability of extracting multi-scale deep features. Specifically, the DBN model was utilized to capture multi-level group-wise temporal features, based on which the individual spatial features were estimated by LASSO regression. Subsequently, a sparse representation method that combines dictionary learning and LASSO regression was utilized to further characterize the group-wise spatial features and individual spatio-temporal features for the purpose of classification. Based on the correspondence between the individual classification features and the group-wise spatial features, a relationship between the decoding capability of classification features and their spatial patterns can be effectively established, which can facilitate the interpretation of neural implications associated with the classification features. Finally, due to its strong generalization capabilities in small sample sizes, SVM was employed for the multi-class classification task.

Our results demonstrated that the proposed framework could successfully classify seven task fMRI signals on a relatively small dataset. Moreover, by taking advantage of DBN in extracting mid-level and high-level features and sparse coding in brain functional network representation (Lv, Jiang, Li, Zhu, Chen, et al., 2015; Ren et al., 2021; Song et al., 2022), our framework could effectively characterize the multi-level spatiotemporal features embedded in multi-task fMRI signals, which provides the bases to identify the interpretable key components for well characterizing and differentiating multi-task signals. Overall, the proposed model can disclose the underlying neural implications of key components with greater classification capacity, offering an effective and interpretable methodology for decoding fMRI data.

Materials and methods

Overview

The framework of our proposed method is illustrated in Figure 1. The pipeline of the proposed framework can divide into four stages: 1) individual data preparation; 2) data preparation for five-fold cross-validation; 3) training and testing process; 4) SVM-based classification and Ratio of activation (ROA) analysis (Fig. 1A). In the data preparation stage, each individual's tfMRI data of seven different tasks were extracted and then spatially concatenated to one signal matrix (the first panel in Fig. 1A). In this work, five-fold cross-validation was performed for model validation, thus the whole dataset was randomly divided into five folds (the second panel in Fig. 1A). In training process, four folds were served as training set, and the tfMRI signal matrices of all the subjects in training set were spatially concatenated to a multi-subject signal matrix. Then, the DBN model was applied to training set to derive the weight matrix W, which served as group-wise temporal features $\boldsymbol{D}^1$. Then, the LASSO regression aims to extract the corresponding loading coefficient $\boldsymbol{\alpha}^1$ based on the defined temporal dictionary $\boldsymbol{D}^1$. In the second stage of our model, the loading coefficient $\boldsymbol{\alpha}^1$ was employed as input to sparse representations (SR) model, where they were decomposed into group-wise dictionaries $\boldsymbol{D}^2$ and loading coefficient $\boldsymbol{\alpha}^2$. In testing process, the individual signal matrix in testing set and the group-wise dictionary $\boldsymbol{D}^1$ obtained during the training phase was utilized as the inputs to the LASSO regression. This yielded the loading coefficients $\boldsymbol{\alpha}^1_{test}$. Subsequently, employing $\boldsymbol{\alpha}^1_{test}$ and the $\boldsymbol{D}^2$ obtained during the training phase, we performed a second LASSO regression to obtain $\boldsymbol{\alpha}^2_{test}$, which were then used as the classification features for the testing subjects (the third panel in Fig. 1A). Note that during the training phase, we utilized the independent training data to learn and train regularization parameters employed for LASSO regression, as well as the group-wise dictionaries $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$, without using any information from the test data. Afterward, to further assess the multi-task fMRI data classification performance of proposed model, the loading coefficient $\boldsymbol{\alpha}^2$ derived from training set was used to train support vector machine (SVM) for classification, where the loading coefficient $\boldsymbol{\alpha}_{test}^2$ derived from testing set was then fed into this trained SVM model to identify the testing set labels (the last panel in Fig. 1A).

Our DBN-SR based framework can also identify the multi-level temporal features, spatial features, and features for multi-task classification (Fig. 1B). Specifically, the DBN model took fMRI time series from training data as input and produced a weight matrix W for each layer respectively, which represent the multi-layer temporal features of group-wise tfMRI signals (the first two panels in Fig. 1B). These multi-layer temporal features W were served as the temporal dictionary $\boldsymbol{D}^1$ and used as input to the LASSO algorithm to regress corresponding loading coefficient $\boldsymbol{\alpha}^1$, which represents individual-level spatial patterns (the third panel in Fig. 1B). Next, the loading coefficient $\boldsymbol{\alpha}^1$ was used as the input of SR stage to derive the common dictionary $\boldsymbol{D}^2$ and the loading coefficient $\boldsymbol{\alpha}^2$, which represent group-wise spatial patterns and features for multi-task classification for each layer, respectively (the last three panels in Fig. 1B).

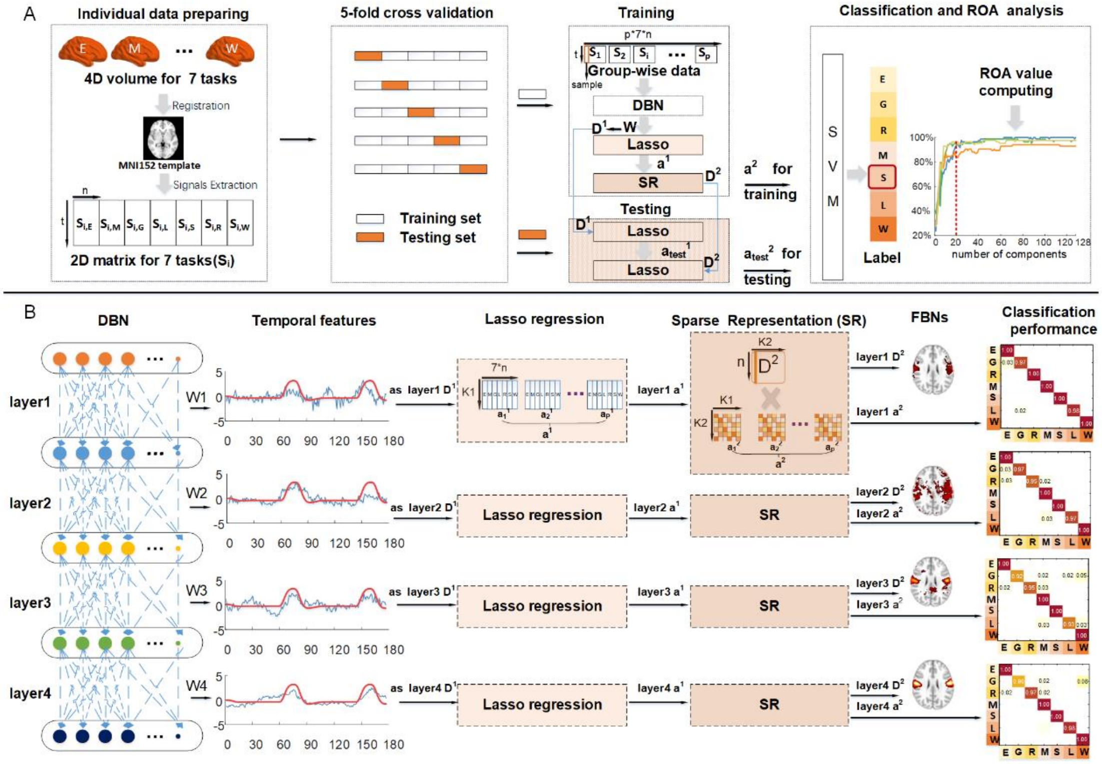


Figure 1. The overview of hybrid deep belief network and sparse representation framework (DBN-SR). (A) The pipeline of multi-task fMRI data classification analysis via the proposed model. The seven capital letters refer to seven different tasks respectively (E: emotion, G: gambling, R: relational, M: motor, L: language, S: social, and W: work memory). (B) The detailed illustration of using DBN and SR model to extract multi-level temporal features, spatial features, and features for classification from multi-task fMRI signals. In the second block, the blue line represents temporal features derived from the weights of DBN, while the red line represents task design paradigms.

Data acquisition and preprocessing

We employed the seven task fMRI data from Q1 release of Human Connectome Project (HCP) in this study (Barch et al., 2013). The details of tfMRI data acquisition and preprocessing pipeline could be referred to our previous study (Song et al., 2022).

Specifically, the seven tasks are emotion, gambling, relational, motor, language, social, and working memory (WM). The number of time points for each task is shown in Table 1. As the tfMRI data consist of different time points, we truncated all tfMRI signals to the same time length (176 frames). In this work, 60 subjects were used from the released dataset

Table1. Details of the condition and frames for seven tasks

TASK	EMOTION	GAMBLING	RELATIONAL	MOTOR	LANGUAGE	SOCIAL	WM
Condition	2	2	2	6	2	2	8
Frames	176	253	232	284	316	274	405

The truncation preprocessing, unavoidably, influences the integrity of task design. For instance, four conditions are excluded from the WM task due to data truncation. Nonetheless, in terms of other tasks, the truncated tfMRI data include not less than one block for all events (sFig. 1).

Data preparation

First, we extracted the whole-brain fMRI signal for each subject using the standard MNI152 template as the mask, resulting in each 2-dimensional matrix. Then the signal matrices of the seven tasks for each subject were spatially concatenated into a large matrix $\boldsymbol{S}_i^1$ ($\boldsymbol{S}_i^1$= [$\boldsymbol{S}_{i,E}^1$, $\boldsymbol{S}_{i,G}^1$, $\boldsymbol{S}_{i,R}^1$, $\boldsymbol{S}_{i,M}^1$, $\boldsymbol{S}_{i,L}^1$, $\boldsymbol{S}_{i,S}^1$, $\boldsymbol{S}_{i,W}^1$] $\in R^{\mathrm{t}\times(\mathrm{n}\times 7)}$, where $\boldsymbol{S}_{i,E}^1 \in R^{\mathrm{t}\times\mathrm{n}}$ had t time points and n voxels. The seven capital letter subscripts refer to seven different tasks respectively (E: emotion, G:

gambling, R: relational, M: motor, L: language, S: social, and W: work memory). TfMRI time series for each voxel were normalized to derive zero mean and unit norm. In this work, five-fold cross-validation scheme was chosen. Thus, 60 subjects were randomly divided into five equal folds. In each iteration, one fold (12 subjects) was taken for testing and the rest four (48 subjects) for training. It is noteworthy that the training and testing sets for each iteration were completely independent. Then, the multi-task fMRI signal matrices of all the subjects in the training set were spatially concatenated to compose a multi-subject fMRI matrix $\boldsymbol{S}^1 = [\boldsymbol{S}_1^1, \boldsymbol{S}_2^1, \ldots, \boldsymbol{S}_p^1] \in R^{\mathrm{t}\times(\mathrm{n}\times 7\times p)}$, where p is the number of training subjects ($p = 48$) (Fig. 1A).

As whole-brain fMRI data generally contain enormous voxels, the group-wise tfMRI signals consisting of multiple tasks and subjects exhibit relatively high dimensionality, inevitably resulting in an overloaded computational burden and memory consumption. To tackle these problems, we randomly selected only 10% of voxels' whole-brain signals for each subject in training stage (Huan Liu 2017; Song et al., 2022). To ensure the uniform distribution of sampled voxels across different brain regions, we employed the Fisher-Yates shuffle algorithm implemented by the "randperm" function in MATLAB, known for generating random permutations with a uniform distribution (Fisher & Yates, 1938). The distribution of the randomly selected 10% voxels across all subjects can be found in the Supplementary Materials (sFig. 6-7).

Deep belief network model-based analysis

In this work, we chose DBN to extract group-wise temporal features based on previous research demonstrating its ability to identify meaningful FBNs (Qiang et al., 2020; Ren et al., 2021). In general, DBN can be regarded as stacked blocks of Restricted Boltzmann Machines (RBM) (G. E. Hinton, Osindero, & Teh, 2006), an energy-based probability generation model that simulates the potential distribution of input data via interactions between visible and hidden variables. While units between visible layer v and hidden layer h are connected by weights, there is no connection within the layer. As a multiple stacked RBM model, the DBN model is designed to learn and train weights for each layer. As described in Asja Fischer (2012) and X. Hu et al. (2018), the energy function of the DBN model adopted to update the weights layer by layer is defined as follows:

$$E(v,h) = \sum b_i v_i - \sum b_j h_j - \sum v_j h_j w_j \tag{1}$$

Where v_i and h_j refer to the activation state of two layers; b_i and b_j represent their bias; w_j indicate the weight between layer i and layer j.

As introduced in the previous section, the tfMRI signals of randomly selected 10% voxels in each individual's whole brain of multi-task in training set were spatially concatenated to generate a multi-subject fMRI matrix for model training, and thus the group-wise tfMRI time series (176 time points) were taken as training samples for the DBN model. In our work, the neural architecture of DBN model was set as 4 layers and 128 neurons experimentally and empirically (see Parameter Selection part). Specifically, the number of visible variables t is the same as the number of time points of fMRI signal (i.e., 176 in our study), and the number of hidden variables $k1$ in each hidden layer represents the number of latent components expressed in fMRI data ($k1$=128). The DBN model was adopted to model group-wise tfMRI matrix $\boldsymbol{S}^1$ to obtain a weight matrix w_j from each layer. The weight matrix of visible layer is represented by $w_1 \epsilon R^{t\times k1}$, and the weight matrix of each hidden layer refers to $w_j \epsilon R^{k1\times k1}$ (j =2,3,4). The multi-layer temporal features W_j in each layer of DBN model can be derived by successive multiplication of the weight matrices on the adjacent layers ($W_j \epsilon R^{t\times k1}$), that is, $W_4 = w_4 * w_3 * w_2 * w_1$, $W_3 = w_3 * w_2 * w_1$, $W_2 = w_2 * w_1$, $W_1 = w_1$. Since each sample input to the DBN model consists of all time points for each voxel, the weights w_j (j =1,2,3,4) across 4 layers represent the temporal features of the input fMRI data at different levels of abstraction. Thus, the successive multiplication of weight matrix W_j (j =1,2,3,4) obtained from each layer of the DBN model represents multi-level temporal features embedded in fMRI signals.

Drawing inspiration from the successful application of LASSO regression for deriving spatial features in previous studies (Haufe et al., 2014; Lee, Jeong, & Ye, 2013), we performed the LASSO regression to derive individual spatial features. Specifically, the multi-layer temporal features W_j derived by the DBN model were normalized and then served as the temporal dictionary $\boldsymbol{D}^1 \epsilon R^{t\times k1}$ (Calhoun et al., 2001; Tibshirani, 2011). Here, as the successive multiplication of weight matrices leads to the larger scale of deeper dictionaries, a normalization procedure ensures reasonable performance of LASSO regression at the same scale. Subsequently, we employed the original individual signal matrix $\boldsymbol{S}_i (i \in 1, 2, \ldots, p)$, along with the temporal dictionary $\boldsymbol{D}^1$ as input to the LASSO algorithm, which produce the corresponding individual loading coefficient $\boldsymbol{\alpha}_i^1$ ($\boldsymbol{\alpha}_i^1 \in R^{k1\times n}$, n=228453). Since $\boldsymbol{D}^1$ incorporates the group-wise temporal features, the resulting individual loading coefficients $\boldsymbol{\alpha}_i^1$ obtained through regression can be considered as spatial sparse representations of each individual's fMRI signals $\boldsymbol{S}_i$ on the common temporal dictionary $\boldsymbol{D}^1$. Consequently, the individual loading coefficients $\boldsymbol{\alpha}_i^1$ represent the individual spatial features. Here, all the loading coefficient matrix derived from LASSO regression refers to $\boldsymbol{\alpha}^1$ ($\boldsymbol{\alpha}^1$=[$\boldsymbol{\alpha}_1^1$, $\boldsymbol{\alpha}_2^1$, …, $\boldsymbol{\alpha}_i^1$, …, $\boldsymbol{\alpha}_p^1$] $\in R^{\mathrm{k1}\times(\mathrm{n}\times 7\times p)}$, $\boldsymbol{\alpha}_i^1$= [$\boldsymbol{\alpha}_{i,E}^1$, $\boldsymbol{\alpha}_{i,G}^1$, $\boldsymbol{\alpha}_{i,R}^1$, $\boldsymbol{\alpha}_{i,M}^1$, $\boldsymbol{\alpha}_{i,L}^1$, $\boldsymbol{\alpha}_{i,S}^1$, $\boldsymbol{\alpha}_{i,W}^1$]$\in R^{\mathrm{k1}\times(\mathrm{n}\times 7)}$.

Similarly, in order to derive the loading coefficient matrix $\boldsymbol{\alpha}_{test}^1$ for testing set of each layer, the group-wise time-series dictionary matrix $\boldsymbol{D}^1$ derived from the training stage was applied to model $\boldsymbol{S}_{test}^1$ to obtain $\boldsymbol{\alpha}_{test}^1$ by resolving a typical l-1 regularized LASSO problem. In this work, the regularization parameter $\lambda 1$ of LASSO regression was set as 0.1 experimentally and empirically.

Sparse Representation model

Although we successfully obtained individual loading coefficient matrices $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}_{test}^1$ through LASSO regression for the training and testing sets, respectively, these features were unsuitable for classification due to their high dimensionality ($\boldsymbol{\alpha}^1 \in R^{k1\times\mathrm{n}}$, $k1$=128, n=228453). Therefore, our next goal was to extract the multi-level group-wise spatial patterns based on the individual spatial patterns, and finally excavate multi-level features for multi-task classification that could distinguish multi-task fMRI signals and reveal the distinctive organization patterns of different task stimulations. Here, we adopted a sparse representation based model, which has already been proven as an effective algorithm in previous research to identify the intrinsic spatial functional patterns and features for multi-task classification from fMRI data (Song et al., 2022; S. Zhang et al., 2016). Specifically, we first aggregated all the loading coefficient matrices $\boldsymbol{\alpha}_i^1$ of all the subjects into one matrix $\boldsymbol{S}^2$ for each layer of the DBN model ($\boldsymbol{S}^2$= [$\boldsymbol{S}_1^2$, $\boldsymbol{S}_2^2$,…,$\boldsymbol{S}_i^2$,…, $\boldsymbol{S}_p^2$] $\in R^{\mathrm{k1}\times(\mathrm{n}\times 7\times \mathrm{p})}$, where $\boldsymbol{S}_i^2$= [$(\boldsymbol{\alpha}_{i,E}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,G}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,R}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,M}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,L}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,S}^1)^{\mathrm{T}}$, $(\boldsymbol{\alpha}_{i,W}^1)^{\mathrm{T}}$] $\in R^{\mathrm{n}\times(7\times \mathrm{k1})}$. Then, $\boldsymbol{S}^2$ would be served as the input for dictionary learning and sparse representation to derive a group-wise spatial dictionary $\boldsymbol{D}^2 \in R^{\mathrm{n}\times\mathrm{k2}}$ and the corresponding loading coefficients $\boldsymbol{\alpha}^2$ for each layer, respectively. Note that $k2$ represents the number of dictionary atoms, which was set as the same value as $k1$ ($k2$=128). Here, $\boldsymbol{\alpha}^{\mathbf{2}}$=[$\boldsymbol{\alpha}_1^2$, $\boldsymbol{\alpha}_2^2$, …, $\boldsymbol{\alpha}_i^2$, …, $\boldsymbol{\alpha}_p^2$]$\in R^{\mathrm{k2}\times(\mathrm{k1}\times 7\times\mathrm{p})}$, where $\boldsymbol{\alpha}_i^2$=[$\boldsymbol{\alpha}_{i,E}^2$, $\boldsymbol{\alpha}_{i,G}^2$, $\boldsymbol{\alpha}_{i,R}^2$, $\boldsymbol{\alpha}_{i,M}^2$, $\boldsymbol{\alpha}_{i,L}^2$, $\boldsymbol{\alpha}_{i,S}^2$, $\boldsymbol{\alpha}_{i,W}^2$]$\in R^{\mathrm{k2}\times\mathrm{k1}\times 7}$. The loss function of sparse representation model yields a sparse resolution constraint on the loading coefficient $\boldsymbol{\alpha}^2$ with an l1 regularization (Eq. (2)), where $\lambda 2$ is a regularization parameter that can balance the regression residual and sparsity level. $\lambda 2$ was set as 0.05.

$$Min\frac{1}{2}\|\boldsymbol{S}^2 - \boldsymbol{D}^2\boldsymbol{\alpha}^2\|_F^2 + \lambda 2\|\boldsymbol{\alpha}^2\|_{1,1} \tag{2}$$

To prevent $\boldsymbol{D}^2$ from arbitrarily large values that cause the trivial solution of the optimization, the columns d_1, d_2, …, d_{k} are restricted by Equation (3).

$$C \triangleq \{\boldsymbol{D}^2 \in R^{\mathrm{t}\times\mathrm{k2}}, s.t. \forall j = 1,\cdots,k2,\ \ d_j^T d_j \leq 1\} \tag{3}$$

As the dictionary $\boldsymbol{D}^2$ was obtained by a sparse representation of $\boldsymbol{\alpha}^{\mathbf{1}}$, which comprise all individual spatial features, the learned dictionary $\boldsymbol{D}^2$consequently represents the group-wise spatial features. Correspondingly, $\boldsymbol{\alpha}_i^2$ was a sparse representation on the common spatial dictionary $\boldsymbol{D}^2$. Given the ability of a sparse representation model to effectively reduce the dimensionality of raw fMRI data while retaining its essential information, the resulting intrinsic features ($\boldsymbol{\alpha}_i^2$) derived from the extraction of common temporal and spatial dictionaries can effectively capture the variations in spatio-temporal patterns of functional brain activity across different tasks. As a result, these intrinsic features were suitable for multi-task classification.

To derive the $\boldsymbol{\alpha}_{test}^2$ of testing set for post-hoc classification analysis, we also leveraged the LASSO regression algorithm for each layer. Specifically, the loading coefficient matrix $\boldsymbol{\alpha}_{test}^1$ was regarded as the input matrix $\boldsymbol{S}_{test}^2$, and the dictionary matrix $\boldsymbol{D}^2$ derived from the training stage was employed to model $\boldsymbol{S}_{test}^{2}$ to learn the loading coefficient $\boldsymbol{\alpha}_{test}^{2}$. All the parameters in testing stage were set the same as in the training stage.

Parameter Selection

The determination of hyperparameters, such as the number of cross-validation folds, the number of layers and neurons of the DBN model, and the regularization parameters of the sparse representation model, was accomplished through a combination of referring to previous studies and learning from the training set, the testing set was not involved in any parameter selection process.

The choice of cross-validation folds is crucial as it offers a trade-off between precision and computational cost for performance estimation (Hansen et al., 2013). Commonly used cross-validation folds in current machine learning experiments often include 2-fold, 5-fold, 10-fold, or the leave-one-out method. In theory, while some studies suggest the 10-fold or leave-one-out method may provide a higher estimated accuracy (Kohavi, 1995), some reveals that 5-fold or 10-fold is the optimal choice for balancing computational cost and accuracy (Hansen et al., 2013). However, due to the need for our framework to combine all individuals within the training set to extract group-wise temporal features during training phase, the computational resource demands of the 10-fold or leave-one-out method are greater. Therefore, we opted for the 5-fold approach. To further validate our selection, we conducted a comparative analysis between the 2-fold and 5-fold to assess the decoding accuracy. The findings revealed that the average decoding rate was slightly lower for the 2-fold compared to the 5-fold, providing additional confirmation of our initial selection. (sTab. 1).

Our selection of a 4-layer, 128-neuron DBN structure was set based on our previous study utilizing the neural architecture search technique (NAS) for recognizing spatio-temporal features from fMRI data (Xu, Ren, Tao, Song, & He, 2022),which effectively determined the optimal structure for DBN model with 3 layers and 120-150 neurons. Therefore, in our study, we defined the number of neurons as 128 and experimented with both 3-layer and 4-layer configurations to extract meaningful task-related temporal features. Specifically, we compared the group-wise temporal features derived from DBN model with 3-layer and 4-layer structures, in terms of their Pearson correlation coefficient (PCC) with task paradigm curve, based on training set (fold 5). The results revealed that the 4-layer DBN outperformed in capturing temporal features, as indicated by the higher PCC values observed in 4-layer structure (Tab. 2). In terms of selecting the number of neurons, we took into consideration computational efficiency. We determined that selecting 128 neurons, a power of two within the desired range of 120-150, would optimize computational speed. Hence, we concluded that the optimal configuration for the DBN model with 128 neurons and 4 layers.

The regularization parameter (λ) plays a crucial role in sparse representation and LASSO regression. Although no golden standard exists for determining the value of λ, previous studies on FBN recognition have experimentally set λ within the range of 0.05 to 0.5 (Fangfei Ge, 2018; Lv, Jiang, Li, Zhu, Chen, et al., 2015; Shu Zhang 2017). In our previous work on task fMRI data classification using a two-stage sparse representation approach, we conducted parameter selection experiments within the range of λ from 0.05 to 0.5 and found that the highest accuracy was achieved when $\lambda1=0.1$ and $\lambda2=0.05$ or 0.1 (Song et al., 2022). Here, $\lambda1$ and $\lambda2$ represent the regularization parameters for the LASSO regression and sparse representation, respectively. Therefore, in this study, we determined the λ1 as 0.1, and systematically changed the setting of the regularization parameter in the sparse representation λ2 (λ2=0.05, 0.1) while evaluating their impact on the obtained group-wise spatial features derived from training set (fold 5). The results showed that when λ2 was set to 0.05, a greater number of FBNs could be identified in the group-wise spatial features $\boldsymbol{D}^2$ by comparison with the general linear model (GLM) -derived activation patterns (Tab. 3). Consequently, we set λ1=0.1 and λ2=0.05 as regularization parameters for LASSO regression and sparse representation stage, respectively. To further validate this, we assessed the classification accuracy on testing dataset using these two different λ2 values (0.05, 0.1) while keeping λ1=0.1 for all 5 folds. The results demonstrated that λ2=0.05 achieved higher accuracy, reconfirming our choice (sTab. 2).

Table 2. Comparison of Pearson correlation coefficient (PCC) for 3-layer structure and 4-layer structure.

Structure	Layer1	Layer2	Layer3	Layer4	Mean±SD
3-layer	0.48±0.12	0.52±0.06	0.50±0.06		0.50±0.08
4-layer	0.55±0.00	0.63±0.01	0.66±0.03	0.71±0.02	0.64±0.02

Table 3. Comparison of the number of identified FBNs cross each layer for different λ2 values.

λ2	Layer1	Layer2	Layer3	Layer4
0.05	15	17	22	45
0.1	12	13	18	27

Identification of multi-level temporal patterns

As mentioned in the "Deep belief network model based analysis" section, W_j of the j-th hidden layer ($j = 1,2,3,4$) represents the temporal features of group-wise tfMRI for respective layer

(Fig. 1B). Here we used PCC as a metric to identify the task-related temporal features (Benesty, Chen, Huang, & Cohen, 2009; Lv, Jiang, Li, Zhu, Chen, et al., 2015). Specifically, we first calculated the task paradigm curves convolved with hemodynamic response function (HRF). Next, we computed the PCC values between the convolved task paradigm curves and the atoms in the group-wise temporal features $\boldsymbol{D}^1$ derived from the DBN model, following standard procedures employed in previous studies (Kay, Rokem, Winawer, Dougherty, & Wandell, 2013; O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). The PCC of the identified temporal features and the task-based stimulus can be defined as Equation (4).

$$P_{corr,\,c} = \mathrm{corr}\ (\boldsymbol{D}_c^1, \mathrm{TASK}) \tag{4}$$

Here, $\boldsymbol{D}_c^1$ refers to the c-th component in temporal features $\boldsymbol{D}^1$ derived from DBN stage (c = 1, $\cdots,k$ 1). TASK represents the task paradigm curves convolved with HRF. Essentially, $P_{corr,\,c}$, measures the temporal similarity between the temporal patterns of $\boldsymbol{D}_c^1$ and the task stimulus. The atoms with the highest PCC value in group-wise temporal features $\boldsymbol{D}^1$ were chosen to represent the multi-layer temporal features.

Identification of multi-level spatial patterns

The multi-level spatial patterns can also be identified in the second stage of sparse representation model. Specifically, the $\boldsymbol{S}_{i,t}^1$ can be factorized into $\boldsymbol{D}^1$ and the loading coefficient $\boldsymbol{\alpha}_{i,t}^1$, which represent the group-wise temporal features and the individual spatial features, respectively. Here, i refers to i-th subjects (i ∈ 1, 2, …, p, and p=48 in this work), t means t kind of task, $t \in \boldsymbol{\Phi} = \{E, G, R, \mathcal{M}, L, \mathcal{S}, W\}$. To further derive the group-wise spatial features, the transposition of $\boldsymbol{\alpha}^1$ could be then decomposed into $\boldsymbol{D}^2$ and $\boldsymbol{\alpha}^2$ as shown in

Equation (5). Since the transpose of $\boldsymbol{\alpha}_{i,t}^{1}$ can be expressed as dictionary $\boldsymbol{D}^{2}$ multiplied by loading coefficient $\boldsymbol{\alpha}_{i,t}^{2}$ (Equation (5)), the relationship between $\boldsymbol{S}_{i,t}^{1}$ and $\boldsymbol{D}^{1}$, $\boldsymbol{D}^{2}$, $\boldsymbol{\alpha}^{2}$ can be deduced as Equation (6) shown, which also consistent with previous studies (Huan Liu 2017; Song et al., 2022).

$$\boldsymbol{S}_{i,t}^{2} = (\boldsymbol{\alpha}_{i,y}^{1})^{T} = \boldsymbol{D}^{2} \times \boldsymbol{\alpha}_{i,t}^{2} \tag{5}$$

$$\boldsymbol{S}_{i,t}^{1} = \boldsymbol{D}^{1} \times \boldsymbol{\alpha}_{i,t}^{1} = \boldsymbol{D}^{1} \times (\boldsymbol{D}^{2} \times \boldsymbol{\alpha}_{i,t}^{2})^{T} \tag{6}$$

Since all subjects share the same group-wise temporal dictionary $\boldsymbol{D}^{1}$, the common dictionary $\boldsymbol{D}^{2}$ contained group-wise spatial patterns, of which atoms could be used to define the FBNs. Thus, the corresponding multi-layer spatial features were derived from the common dictionary $\boldsymbol{D}^{2}$ for each layer of the proposed framework (the fourth and fifth panels in Fig. 1B).

We then identified the spatial correlation coefficient (SCC) to quantify the similarity between spatial patterns obtained from the proposed framework and the GLM -derived activation patterns. Specifically, the GLM-based analysis was performed individually, followed by group-wisely analysis using FSL FEAT (http://www.fmrib.ox.ac.uk/fsl/feat5/index.html). The group-level GLM-based results were employed for comparison. More details of GLM analysis are available in previous literature (Lv, Jiang, Li, Zhu, Zhang, et al., 2015). The SCC is defined in Equation (7) (Ben J. Harrison, 2008; Zuo et al., 2010):

$$\mathbf{R}\ (\boldsymbol{X}, \boldsymbol{T}) = \frac{\Sigma_{p=1}^{n}(X_p-\bar{X})(T_p-\bar{T})}{\sqrt{\Sigma_{p=1}^{n}(X_p-\bar{X})^2 \cdot \Sigma_{p=1}^{n}(T_p-\bar{T})^2}} \tag{7}$$

where $\boldsymbol{X}$ is the spatial functional network derived by the proposed framework, $\boldsymbol{T}$ represents the GLM-derived activation template, and n refers to the number of voxels of whole brain.

SVM-based classification method

To further classify multi-task fMRI signals, we performed five-fold cross-validation to evaluate the classification performance of the proposed framework. As the linear SVM has optimization and generalization capability in limited sample sizes, as well as its proven effectiveness in multi-class classification (Chang & Lin, 2011b; Jang et al., 2017), we conducted multi-task classification analysis based on linear SVM classifier, which was established by the LIBSVM toolbox (Chang & Lin, 2011a). For each layer, as the loading coefficient $\boldsymbol{\alpha}^2$ contains both temporal and spatial features embedded in fMRI signals, we first trained the SVM classifier using $\boldsymbol{\alpha}^2$ derived from training set, and then evaluated the classification performance by feeding the $\boldsymbol{\alpha}_{test}^2$ of testing set into the trained SVM model. Based on the true label of seven tasks for each loading coefficient $\boldsymbol{\alpha}_{test}^2$, the classification accuracy of each layer in each fold was defined as the percentage of correctly predicted samples. The final classification accuracy for each layer is the average of five folds for seven tasks. We then calculated the specificity of each fold for each layer, and the final specificity for each layer is the average of the five folds.

ROA-based analysis

The further goal aimed at uncovering discriminative functional components for multi-task classification. Inspired by the successful use of the Ratio of activation (ROA) in identifying discriminative components for decoding resting state fMRI (rsfMRI) and tfMRI (S. Zhang et al., 2016), we raised a novel ROA metric to identify the key components for seven-task classification. The ROA of the j-th row in loading coefficients $\boldsymbol{\alpha}^2$ could be defined as follows:

$$N_t = |\boldsymbol{\alpha}^2(j,k)|_0, kth\ column\ belongs\ to\ task(t)$$

$$\mathrm{ROA}_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(N_t - \overline{N_t})^2} \tag{8}$$

In Equation (8), $\boldsymbol{\alpha}^2$ represent all the individual spatio-temporal features, $\boldsymbol{\alpha}^2$= [$\boldsymbol{\alpha}_1^2$, $\boldsymbol{\alpha}_2^2$, …, $\boldsymbol{\alpha}_i^2$, …, $\boldsymbol{\alpha}_p^2$]$\in R^{\mathrm{k2}\times(\mathrm{k1}\times 7\times \mathrm{p})}$ (k1= k2=128, p=48). i refers to i-th subject ($i \in$1, 2, …, p). t represents task index (t$\in$1, 2, …, 7), and T represents the number of task paradigms (i.e., 7 in our work). Task (t) represents each of the seven different tasks. N_t represents the activation level for each task, and $\overline{N_t}$ represents the average of N_t (t = 1, ⋯,7). Here, the activation level N_t was defined by counting the number of non-zero entries marked as each task in the corresponding each row vector of $\boldsymbol{\alpha}^2$ (t$\in$1, 2, …, 7). As $\boldsymbol{\alpha}^2$ is a sparse matrix, the task with a higher count of nonzero elements in the row vectors of $\boldsymbol{\alpha}^2$ is deemed to be more "active". Therefore, N_t represents each task's activation level in the row vectors of $\boldsymbol{\alpha}^2$. The ROA was calculated by counting the standard deviation of N_t across the seven tasks. A larger ROA value (i.e., larger standard deviation) indicates greater differences in activity levels across the seven tfMRI signals, which were more discriminative for multi-task classification.

To validate that the components of higher ROA values capture greater capacity in classifying the multi-task fMRI signals, an experiment was designed as illustrated below. After sorting the ROA values for all components (i.e., rows in loading coefficients $\boldsymbol{\alpha}^2$) from highest to lowest, we iteratively adopted more rows sorted by their ROA values in $\boldsymbol{\alpha}^2$ as feature inputs for training the SVM classifier, that is, the components with higher ROA values were used preferentially for training. Afterwards, the corresponding components of $\boldsymbol{\alpha}_{test}^2$ from testing set were entered into the trained SVM model to evaluate the classification accuracy. Specifically, to define the key components with greater capacity for multi-task classification in each layer, we have repeated this ROA analysis using $\boldsymbol{\alpha}^2$ derived from each layer of proposed model. Here we applied the same classification scheme described in the previous section "SVM-based classification method".

After establishing the ROA metric for the classification features $\boldsymbol{\alpha}^2$, our subsequent objective is to elucidate the neural implications of these classification features. Given that each row of $\boldsymbol{\alpha}^2$ corresponds to each column of $\boldsymbol{D}^2$ (i.e., each atom in $\boldsymbol{D}^2$), and these atoms can be mapped back to brain space, we thus established a relationship between the brain activations derived from the atoms in $\boldsymbol{D}^2$ and the ROA values of the row vectors of $\boldsymbol{\alpha}^2$. This connection allows us to interpret neural implications of classification features.

Result

Classification performance of multi-task fMRI signals

By applying the proposed DBN-SR framework to multi-task fMRI data using five-fold cross-validation strategy, our results reveal that the fMRI data of seven tasks can be accurately classified. In detail, the classification accuracy for five-fold ranges from 92.86% to 100%, with an average accuracy of 97.86%±3.42% (Mean ± SD) in the layer 4 (Fig. 2A), which demonstrated the proposed framework can effectively uncover the inherent differences in composition patterns of multi-task fMRI signals.

We also explored the classification performance based on features derived from each layer of the proposed framework (Fig. 2). The trend of the classification accuracy curves for five folds is relatively steady, with an average accuracy of 98.15%±0.90% (Mean±SD) (Fig. 2A). Moreover, the average accuracies across five-fold from layer1 to layer4 are 99.29%, 98.33%, 97.14%, and 97.86%, respectively. We depicted confusion matrices for each layer to represent the average classification accuracy of the seven tasks, as shown in Figure 2b. The results indicate that all the average classification accuracies for seven tasks across five-fold are greater than 95% in each layer, except for three major confusions, that is, gambling task in layer 3 and layer 4, relational task in layer 2 and layer 3, and language task in layer 3 (Fig. 2B). In addition, the specificity of classification results of the first two layers is slightly higher than that of the deeper two layers (Fig. 2C). Overall, the classification performance of the shallower layers is relatively better than that of the deeper layers.

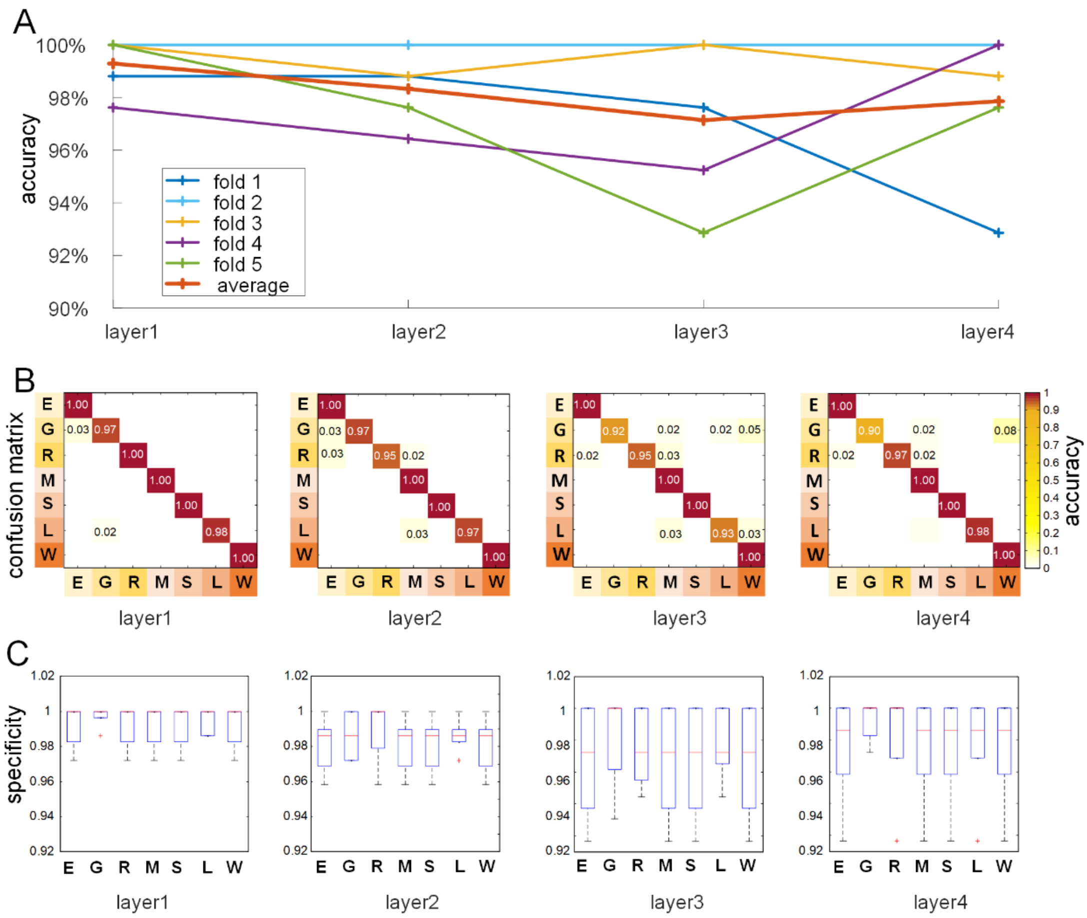


Figure 2. Classification performance. (A) The classification accuracy of five-fold in each layer. (B) The average confusion matrices of five-fold cross-validation on the seven tasks. (C) The

average specificity of five-fold cross-validation classification on the seven tasks.

Identified multi-level temporal and spatial patterns of multi-task fMRI signals

Multi-level temporal patterns

Our DBN-SR based framework can effectively identify the temporal patterns of multi-task fMRI signals at multi-scale (Fig. 3). In each layer, we quantitatively compared the PCC of the identified temporal features and each task-based stimulus. Those atoms with the highest PCC value in temporal dictionary $\boldsymbol{D}^1$ were chosen to represent the task-related temporal patterns. We randomly select one training fold as an example to show the representative temporal patterns for each layer (fold 5) (Fig. 3). The average PCC values of seven tasks for all 5-fold can be found in Supplemental Table 6.

The overall multi-level temporal patterns are relatively consistent with the task design paradigms. Specifically, the average PCC of seven tasks from layer1 to layer4 is 0.55±0.12, 0.61±0.03, 0.65±0.07, and 0.71±0.08 (Mean ± SD), respectively, where the highest correlation is observed in layer4 (Fig. 3). Intriguingly, there exist gradient in the resolution of temporal patterns derived from different layers. In the shallow layer, all the identified temporal patterns are mixed with many random noises, resulting in a relatively poor correlation with task paradigms. In comparison, in the deeper layer, the temporal patterns are smoother and more consistent with the original task design curves, indicating that DBN-SR model can filter noises in each layer while keeping useful information of brain activities, which agrees with the former research (H. Huang et al., 2018; Wei Zhang, 2020).

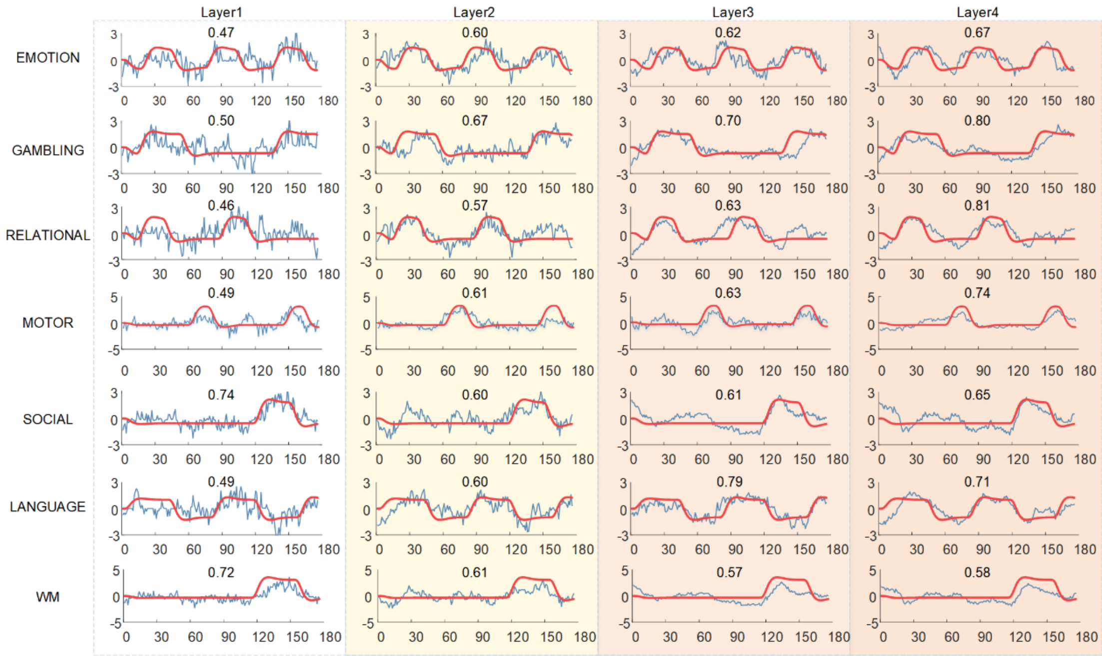


Figure 3. Comparison of group-wise temporal patterns for seven tasks across different layers, including the identified temporal features (blue lines) and the task paradigms (red lines). The quantitative similarities (PCC) of identified temporal features with task paradigms are also provided. The y-axis represents the stimulus response amplitude, while the x-axis represents time point. The background colors represent different layers of our DBN-SR model. The lighter colors represent shallower layers, while the darker colors represent deeper layers.

Multi-level spatial patterns

Our framework can also effectively identify the spatial patterns from different layers. The most predominant spatial patterns identified by the proposed framework are the task-evoked FBNs, including emotion, gambling, relational, motor, social, language, and working memory. In each layer, we quantitatively compared the SCC of the identified spatial patterns and the GLM-derived activation patterns. Those atoms with the highest SCC value in spatial dictionaries $\boldsymbol{D}^2$ were chosen to represent the spatial pattern. We randomly selected one training fold to illustrate the representative FBNs for each layer (Fig. 4).

Overall, the spatial patterns are generally consistent with the GLM-derived activation patterns, with increasingly precise resolution from shallow to deep layers. Quantitatively, the average SCC of seven tasks from layer1 to layer4 is 0.36±0.20, 0.26±0.11, 0.40±0.12, and 0.48±0.12 (Mean ± SD), respectively, where the highest SCC is observed in layer 4 (Fig. 4). Intriguingly, there exist distinct differences among spatial patterns derived from different layers. The spatial patterns across layers show a trend of increasing consistency with the GLM-derived activation patterns, and are more compact in deeper layers for most tasks. Meanwhile, more FBNs can be found in the deeper layers compared with shallow layer. For example, some FBNs cannot be identified in the first three layers, such as FBNs related to gambling and relational tasks (Fig. 4).

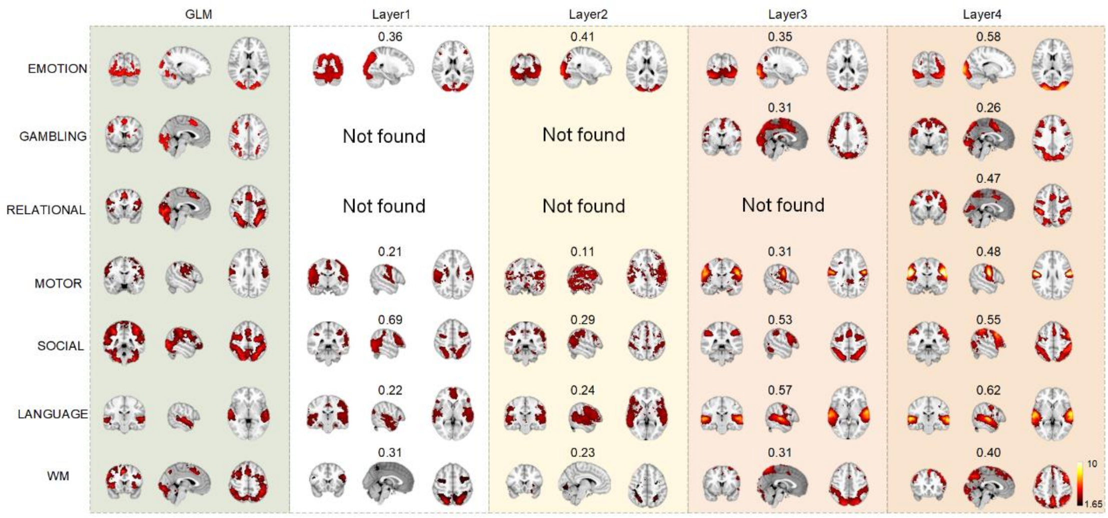


Figure 4.

Comparison of group-wise spatial patterns for seven tasks across different layers. The spatial correlation coefficient (SCC) between each identified spatial pattern and GLM-derived activation pattern is labeled on top of each brain map.

Apart from FBNs, the proposed framework can also effectively detect various artifact-related components. Specifically, the atoms in spatial dictionary $\boldsymbol{D}^2$ can represent the group-wise spatial features and can be mapped back to the 3D brain volume. Subsequently, we manually inspected whether spatial map matched the known types of artifacts based on previous study (Salimi-Khorshidi et al., 2014). Through this process, we found several artifact-related components, including movement-related, cardiac-related, sagittal sinus, susceptibility-motion, white-matter, and MRI acquisition/reconstruction related (Fig. 5).

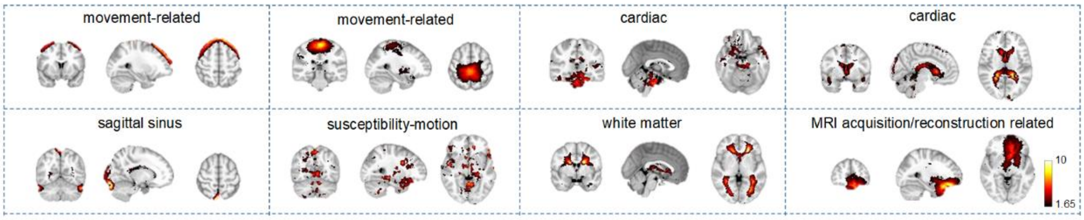


Figure 5. Identified artifact components, including movement-related, cardiac-related, sagittal sinus, susceptibility-motion, white-matter, and MRI acquisition/reconstruction related.

Overall, our effective DBN-SR model is capable of characterizing the multi-level spatiotemporal features of brain function. The quantitative analysis further demonstrates that, in deeper layer, the representative temporal features correspond well with task design curves, and the spatial features are relatively more consistent with the GLM-derived activation. In addition to task-evoked functional components, our framework could also effectively identify artifact components from group-wise multi-task fMRI data, laying the groundwork for further research into the functional role of these components in multi-task classification.

Identification of discriminative features by ROA analysis

As depicted in the “ROA-based analysis” section, we first computed the ROA index by sorting the ROA values of all the components in loading coefficients $\boldsymbol{\alpha}^2$ of the training set, then, in order to evaluate the classification performance, the corresponding components in the loading coefficient $\boldsymbol{\alpha}_{test}^2$ of testing set were fed sequentially into the trained SVM classifier according to the ROA index. Here, the classification results of each layer on one randomly selected testing fold dataset (fold 5) using different number of components, sorted by their ROA values, are illustrated in Fig. 6A. While the number of components increases from 1 to 20, the accuracy curves of four layers grow monotonically, and the average accuracy of all curves rises to 91.96%. When more than twenty components are included for classification, the accuracy curves of four layers exhibit a plateau with accuracies reaching close to 100%, indicating that the additional components with lower ROA values contribute less to the successful classification of multi-task signals. Thus, the top twenty components with higher ROA values can be regarded as key components for the classification task to some extent. Generally, our method can effectively disclose the key components with great classification capacity. In addition, the findings are consistent across different testing folds, hence the additional results of the other four folds are included in the Supplementary Materials (sFig2-5).

To further investigate the neural implications of key components with greater classification capacity, we inspected the spatial patterns of the top twenty key components identified by ROA analysis in each layer. By further analyzing the composition of the twenty key components in each layer, we found that these key atoms are either FBNs or artifact-related components, which were identified by visually examining their spatial patterns with established templates and further calculating their SCC with GLM-derived activation maps.

Intriguingly, our results show that the top twenty key components in the four layers are largely composed of artifacts, while the proportion of FBNs in key components is small as a whole. On the other hand, the proportion of FBNs is relatively higher in deeper layers compared to shallower layers (Fig. 6B). This conclusion aligns with the findings when using the top 40 components as key components (sFig. 8).

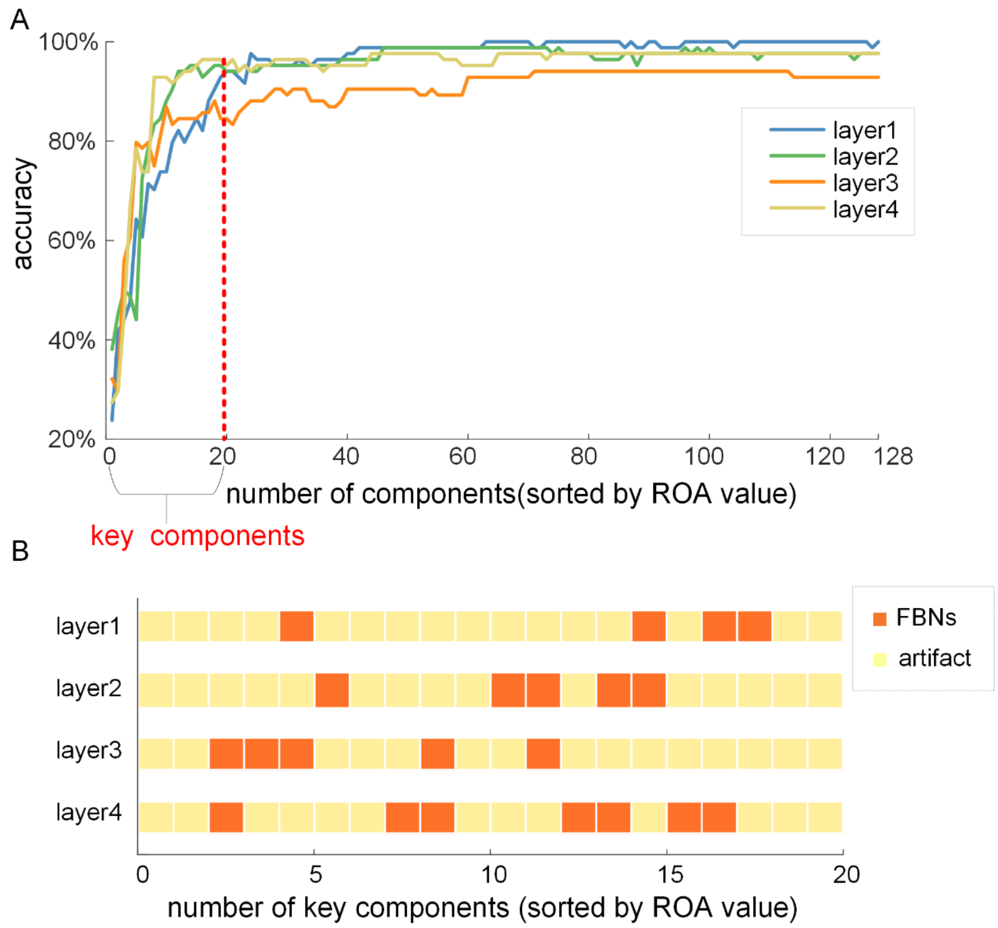


Figure 6. ROA classification results in each layer (fold 5). (A) Classification accuracy for SVM-based classification of four layers using the different number of components sorted by their ROA values. (B) The composition of twenty key components sorted by ROA value across each layer.

Discussion

In this study, we proposed a hybrid spatio-temporal deep belief network and sparse representation framework to decode multi-task fMRI signals on a relatively small cohort dataset. Our framework could classify fMRI signals of seven tasks with high accuracy and detect multi-level temporal patterns and FBNs, suggesting the effectiveness of the proposed method. In addition, our framework can reveal key components including artifact components and functional brain networks in multi-task classification and uncover their underlying neurological implication.

Our proposed framework is composed of several elements, including DBN model, LASSO regression, sparse representation, and SVM classifier, resulting in a relatively complex structure. Nevertheless, our framework achieved a relatively higher classification accuracy in comparison to prior research that also conducted classification of 7 task states on the HCP dataset (X. Huang, Xiao, & Wu, 2021; Wang et al., 2020), while also yielding interpretable classification components. Specifically, Wang et al. (2020) reported two standard machine learning algorithms, namely MVPA-SVM and DNN, and X. Huang et al. (2021) proposed a novel framework (CRNN) incorporating multiple modules such as CNN, recurrent neural network (RNN), and attention mechanism. The average accuracy of our framework (98.15%) is much higher than that of MVPA-SVM (69.2%) and comparable to the accuracies of DNN-based model (93.7%) and CRNN-based model (94.31%) (X. Huang et al., 2021; Wang et al., 2020). Additionally, the neuroscientific implications of their results remain elusive. In conclusion, our proposed model achieved higher decoding accuracy than these models, while also providing a more comprehensive and interpretable methodology for decoding fMRI data.

Furthermore, our model unveils multi-level temporal and spatial patterns, demonstrating a resolution gradient spanning from shallow to deep layers. Specifically, in the deeper layers, the identified temporal features are better correlated to the original task paradigm curves.

Meanwhile, more diverse FBNs can be detected and the spatial features show more consistency with the GLM-derived activation patterns, in deeper layers.

Intriguingly, although more higher-order FBNs can be detected in deeper layers, the classification accuracy using features for multi-task classification derived from deeper layers is lower than that of shallower layers, indicating that these higher-order FBNs are not much helpful for multi-task classification. To validate this observation, we specifically selected only FBNs components from all available components across all five folds for multi-task classification, resulting in an average accuracy of 97.08%±2.14% (Mean±SD), slightly lower than the classification rate obtained using all components (98.15%±0.90%) (sTab. 3). The possible reason is that the FBNs evoked by different cognitive tasks may have co-activated brain regions, thus the FBNs components alone may not fully reveal the potential fundamental differences in functional composition patterns of multi-task fMRI data. On the other hand, ROA-based analyses indicate that artifact components occupy higher proportion of key components for multi-task classification in shallower layers than that in deeper layers, along with higher classification accuracy and specificity in the shallower layers. These findings suggest that the artifact components play an important role in multi-task fMRI signal classification, which is also consistent with previous research, where the artifact components of the EEG signal are significantly more informative than brain activity concerning classification accuracy (McDermott et al., 2021).

While our study provides novel insight into the core functional components in decoding multi-task fMRI signals, it is important to note that there are three limitations. The first limitation is the manual setting of parameters for DBN and sparse representation framework, mainly including the number of neuron nodes and layers in DBN and the sparsity penalty parameter of SR. Thus, automatic optimization of model parameters is one of the future research directions. The second limitation stems from our inability to detect FBNs related to gambling and relational tasks within the first two to three layers of the DBN-SR framework. This could be attributed to more noise present in the group-wise temporal features $\boldsymbol{D}^1$ extracted at lower levels (Fig. 1). Additionally, LASSO regression may not be well-suited for handling noisy shallow features, thus making it challenging for LASSO regression to accurately capture the underlying spatial patterns. To address this limitation, future studies could explore alternative regression approaches that are better suited for handling noisy shallow features, thereby improving the accurate acquisition of the underlying spatial patterns. The third limitation is that our study employed a relatively small dataset, consisting of 60 individuals out of 68 from HCP Q1 dataset. To assess the robustness of our model, we included the remaining 8 individuals from the same dataset as a hold-out dataset, 6 of which do not have complete data for all 7 tasks (sTab. 4). However, this does not affect their suitability as an independent lock box dataset to test the performance of our trained model. The results revealed that the average decoding accuracy for these 8 individuals (96.43%) was comparable to the 5-fold cross-validation accuracy of the 60 individuals (sTab. 5), suggesting the robustness of our model. Nonetheless, we acknowledge that a larger dataset would lend further support to our findings. In future work, we aim to apply our model to more extensive or multicenter datasets to evaluate its generalizability and robustness.

Overall, with the superiority of interpretability and effectiveness of DBN-SR model on small datasets, our framework could potentially be useful to differentiate abnormal brain function in clinical research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant. No. 62006187), the Youth Innovation Team Foundation of Education Department of Shaanxi Province Government (Grant. No. 21JP119), the China Postdoctoral Science Foundation Funded Project (Grant No. 2021M702650), the National Natural Science Foundation of China (Grant. No. 61971350), the National Natural Science Foundation of China (Grant. No. 12271434), Natural Science Basic Research Program of Shaanxi (Grant. No. 2023-JC-JQ-57), and the Key Research and Development Program Project of Shaanxi Province (Grant. No. 2020SF-036). We thank the Human Connectome Project for providing Quarter 1 (Q1) Dataset (https://www.humanconnectome.org/study/hcp-young-adult/document/q1-data-release).

Reference

Asja Fischer, C. I. (2012). An Introduction to Restricted Boltzmann Machines. Paper presented at the Iberoamerican Congress on Pattern Recognition, Berlin.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., . . . Consortium, W. U.-M. H. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. Neuroimage, 80, 169-189. doi:10.1016/j.neuroimage.2013.05.033

Ben J. Harrison, J. P., Marina Lo´ pez-Sola, Rosa Herna´ ndez-Ribas, Joan Deus, Hector Ortiz, Carles Soriano-Mas, Murat Yu¨ cel, Christos Pantelis, and Narcı´s Cardoner. (2008). Consistency and functional specialization in the default mode brain network. PNAS, 105, 9781–9786.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4): Springer.

Bengio, Y., Courville, A. C., & Vincent, P. (2012). Unsupervised feature learning and deep learning: A review and new perspectives. CoRR, abs/1206.5538, 1(2665), 2012.

Bo Liu, Y. W., Yu Zhang, Qiang Yang. (2017, August). Deep Neural Networks for High Dimension, Low Sample Size Data. Paper presented at the IJCAI, Melbourne.

Calhoun, V. D., Adali, T., McGinty, V. B., Pekar, J. J., Watson, T. D., & Pearlson, G. D. (2001). fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. Neuroimage, 14(5), 1080-1088. doi:10.1006/nimg.2001.0921

Chang, C.-C., & Lin, C.-J. (2011a). Libsvm. ACM Transactions on Intelligent Systems and Technology, 2(3), 1-27. doi:10.1145/1961189.1961199

Chang, C.-C., & Lin, C.-J. (2011b). LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 1-27.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., . . . Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage, 28(3), 663-668. doi:10.1016/j.neuroimage.2005.08.009

Dong, Q. (2020). Modeling Hierarchical Brain Networks via Volumetric Sparse Deep Belief Network (VSDBN). Computerized Medical Imaging and Graphics.

Fangfei Ge, J. L., Xintao Hu , Lei Guo , Junwei Han , Shijie Zhao, Tianming Liu (2018, April 4-7). Exploring intrinsic networks and their interactions using group wise temporal sparse coding.

Paper presented at the International Symposium on Biomedical Imaging (ISBI 2018), Washington, D.C., USA.

Fisher, R. A., & Yates, F. (1938). Statistical tables for biological, agricultural aad medical research. Statistical tables for biological, agricultural aad medical research.

Friston, K. J. (2009). Modalities, Modes, and Models in Functional Neuroimaging. SCIENCE, 326, 399-403.

Hansen, K., Montavon, G., Biegler, F., Fazli, S., Rupp, M., Scheffler, M., . . . Muller, K. R. (2013). Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. J Chem Theory Comput, 9(8), 3404-3419. doi:10.1021/ct400195d

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage, 87, 96-110.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. Nat Rev Neurosci, 7(7), 523-534. doi:10.1038/nrn1931

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Comput, 18(7), 1527-1554. doi:10.1162/neco.2006.18.7.1527

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. Parallel distributed processing: Explorations in the microstructure of cognition, 1(282-317), 2.

Hu, J., Kuang, Y., Liao, B., Cao, L., Dong, S., & Li, P. (2019). A Multichannel 2D Convolutional Neural Network Model for Task-Evoked fMRI Data Classification. Comput Intell Neurosci, 2019, 5065214. doi:10.1155/2019/5065214

Hu, X., Huang, H., Peng, B., Han, J., Liu, N., Lv, J., . . . Liu, T. (2018). Latent source mining in FMRI

via restricted Boltzmann machine. Hum Brain Mapp, 39(6), 2368-2380. doi:10.1002/hbm.24005
Huan Liu , M. Z., Xintao Hu , Yudan Ren , Shu Zhang , Junwei Han , Lei Guo , Tianming Liu (2017). Fmri data classification based on hybrid temporal and spatial sparse representation. Paper presented at the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia.
Huang, H., Hu, X., Zhao, Y., Makkie, M., Dong, Q., Zhao, S., . . . Liu, T. (2018). Modeling Task fMRI Data Via Deep Convolutional Autoencoder. IEEE Trans Med Imaging, 37(7), 1551-1561. doi:10.1109/TMI.2017.2715285
Huang, X., Xiao, J., & Wu, C. (2021). Design of Deep Learning Model for Task-Evoked fMRI Data Classification. Comput Intell Neurosci, 2021, 6660866. doi:10.1155/2021/6660866
Jang, H., Plis, S. M., Calhoun, V. D., & Lee, J. H. (2017). Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. Neuroimage, 145(Pt B), 314-328. doi:10.1016/j.neuroimage.2016.04.003
Kay, K., Rokem, A., Winawer, J., Dougherty, R., & Wandell, B. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Frontiers in neuroscience, 247.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Paper presented at the Ijcai.
Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage, 38(4), 649-662.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. doi:10.1038/nature14539

Lee, J., Jeong, Y., & Ye, J. C. (2013). Group sparse dictionary learning and inference for resting-state fMRI analysis of Alzheimer's disease. Paper presented at the 2013 IEEE 10th International Symposium on Biomedical Imaging.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Med Image Anal, 42, 60-88. doi:10.1016/j.media.2017.07.005

Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-task deep neural networks for natural language understanding. arXiv preprint arXiv:1901.11504.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869-878.

Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., . . . Liu, T. (2015). Sparse representation of whole-brain fMRI signals for identification of functional networks. Med Image Anal, 20(1), 112-134. doi:10.1016/j.media.2014.10.011

Lv, J., Jiang, X., Li, X., Zhu, D., Zhang, S., Zhao, S., . . . Liu, T. (2015). Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. IEEE Trans Biomed Eng, 62(4), 1120-1131. doi:10.1109/TBME.2014.2369495

McDermott, E. J., Raggam, P., Kirsch, S., Belardinelli, P., Ziemann, U., & Zrenner, C. (2021). Artifacts in EEG-Based BCI Therapies: Friend or Foe? Sensors (Basel), 22(1). doi:10.3390/s22010096

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of big data, 2(1), 1-21.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. Social cognitive and affective neuroscience, 7(5), 604-609.

Qiang, N., Dong, Q., Zhang, W., Ge, B., Ge, F., Liang, H., . . . Liu, T. (2020). Modeling task-based fMRI data via deep belief network with neural architecture search. Comput Med Imaging Graph, 83, 101747. doi:10.1016/j.compmedimag.2020.101747

Rashid, M., Singh, H., & Goyal, V. (2020). The use of machine learning and deep learning algorithms in functional magnetic resonance imaging—a systematic review. Expert Systems, 37(6), e12644. doi:10-1111

Ren, Y., Xu, S., Tao, Z., Song, L., & He, X. (2021). Hierarchical Spatio-Temporal Modeling of Naturalistic Functional Magnetic Resonance Imaging Signals via Two-Stage Deep Belief Network With Neural Architecture Search. Front Neurosci, 15, 794955. doi:10.3389/fnins.2021.794955

Rubin, T. N., Koyejo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., & Yarkoni, T. (2017). Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLoS Comput Biol, 13(10), e1005649. doi:10.1371/journal.pcbi.1005649

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage, 90, 449-468. doi:10.1016/j.neuroimage.2013.11.046

Shu Zhang , X. L., Lei Guo , Tianming Liu. (2017, 18-21 April). Exploring human brain activation via nested sparse coding and functional operators. Paper presented at the International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia.

Song, L., Ren, Y., Hou, Y., He, X., & Liu, H. (2022). Multitask fMRI Data Classification via Group-Wise

Hybrid Temporal and Spatial Sparse Representations. eNeuro, 9(3). doi:10.1523/ENEURO.0478-21.2022

Sotetsu Koyamadaa, b., Yumi Shikauchia,b, Ken Nakaea, Masanori Koyamaa, Shin Ishii. (2015). Deep learning of fMRI big data: a novel approach to subject-transfer decoding. arXiv preprint arXiv.

Stanislas Dehaene, G. L. C. H., Laurent Cohen, Jean-Baptiste Poline, Pierre-François van de Moortele and Denis Le Bihan. (1998). Inferring behavior from functional brain images.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. Royal Statistical Society, 73, 273-282.

Varoquaux, G., & Thirion, B. (2014). How machine learning is shaping cognitive neuroimaging. GigaScience, 3(1), 1-7. doi:10.1186

Vieira, S., Pinaya, W. H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci Biobehav Rev, 74(Pt A), 58-75. doi:10.1016/j.neubiorev.2017.01.002

Wang, X., Liang, X., Jiang, Z., Nguchu, B. A., Zhou, Y., Wang, Y., . . . Qiu, B. (2020). Decoding and mapping task states of the human brain via deep learning. Hum Brain Mapp, 41(6), 1505-1519. doi:10.1002/hbm.24891

Wei Zhang, S. Z., Xintao Hu,2, Qinglin Dong,Heng Huang,Shu Zhang, Yu Zhao, Haixing Dai, Fangfei Ge, Lei Guo and Tianming Liu. (2020). Hierarchical Organization of Functional Brain Networks Revealed by Hybrid Spatiotemporal Deep Learning. Brain Connectivity, 10. doi:10.1089/brain.2019.0701

Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., & Han, W. (2018). Deep Learning Methods to Process fMRI Data and Their Application in the Diagnosis of Cognitive Impairment: A Brief Overview and Our Opinion. Front Neuroinform, 12, 23. doi:10.3389/fninf.2018.00023

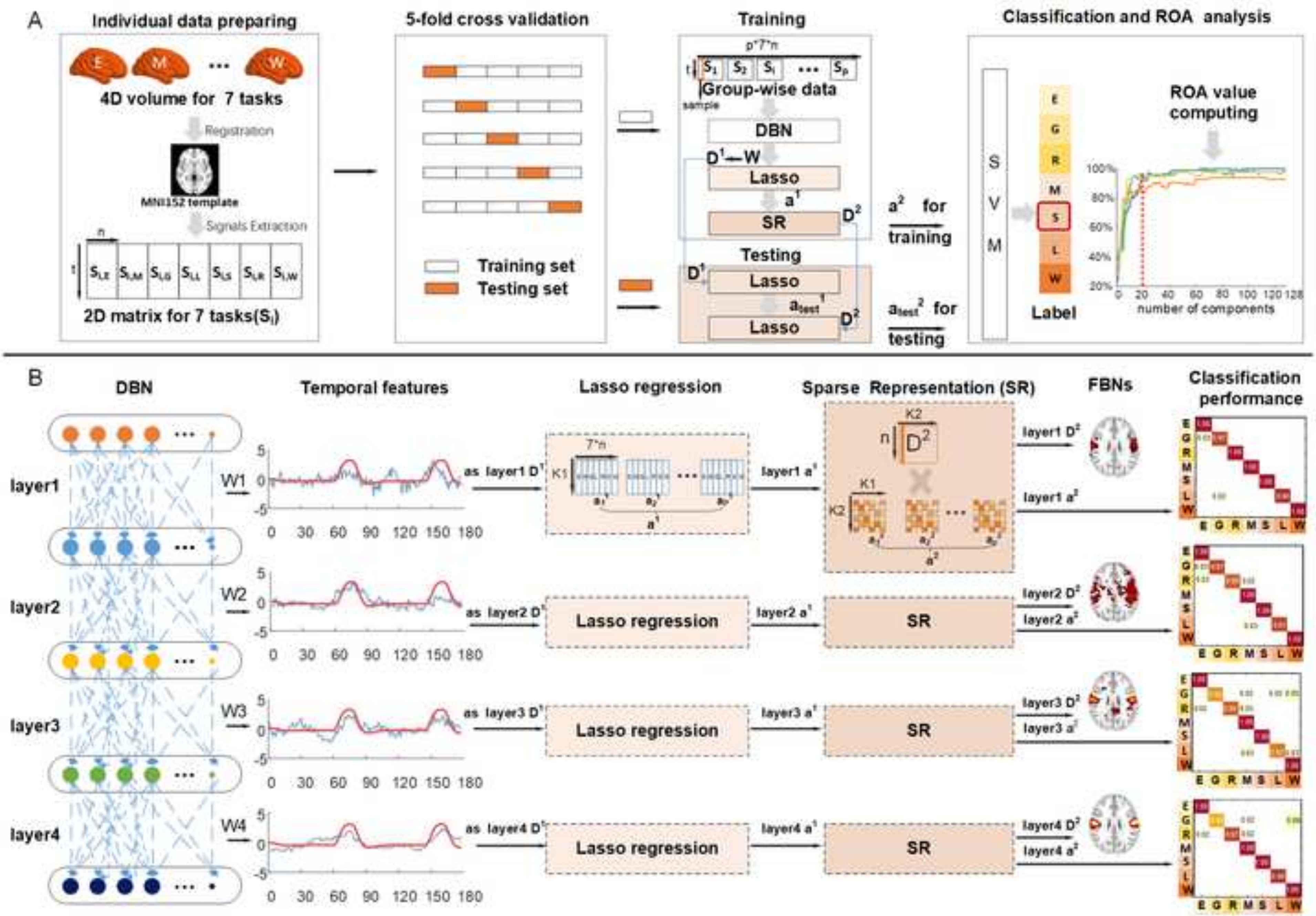
Xu, S., Ren, Y., Tao, Z., Song, L., & He, X. (2022). Hierarchical Individual Naturalistic Functional Brain Networks with Group Consistency uncovered by a Two-Stage NAS-Volumetric Sparse DBN Framework. eNeuro, 9(5). doi:10.1523/ENEURO.0200-22.2022

Zhang, S., Li, X., Lv, J., Jiang, X., Guo, L., & Liu, T. (2016). Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations. Brain Imaging Behav, 10(1), 21-32. doi:10.1007/s11682-015-9359-7

Zhang, Y., Tetrel, L., Thirion, B., & Bellec, P. (2021). Functional annotation of human cognitive states using deep graph convolution. Neuroimage, 231, 117847. doi:10.1016/j.neuroimage.2021.117847

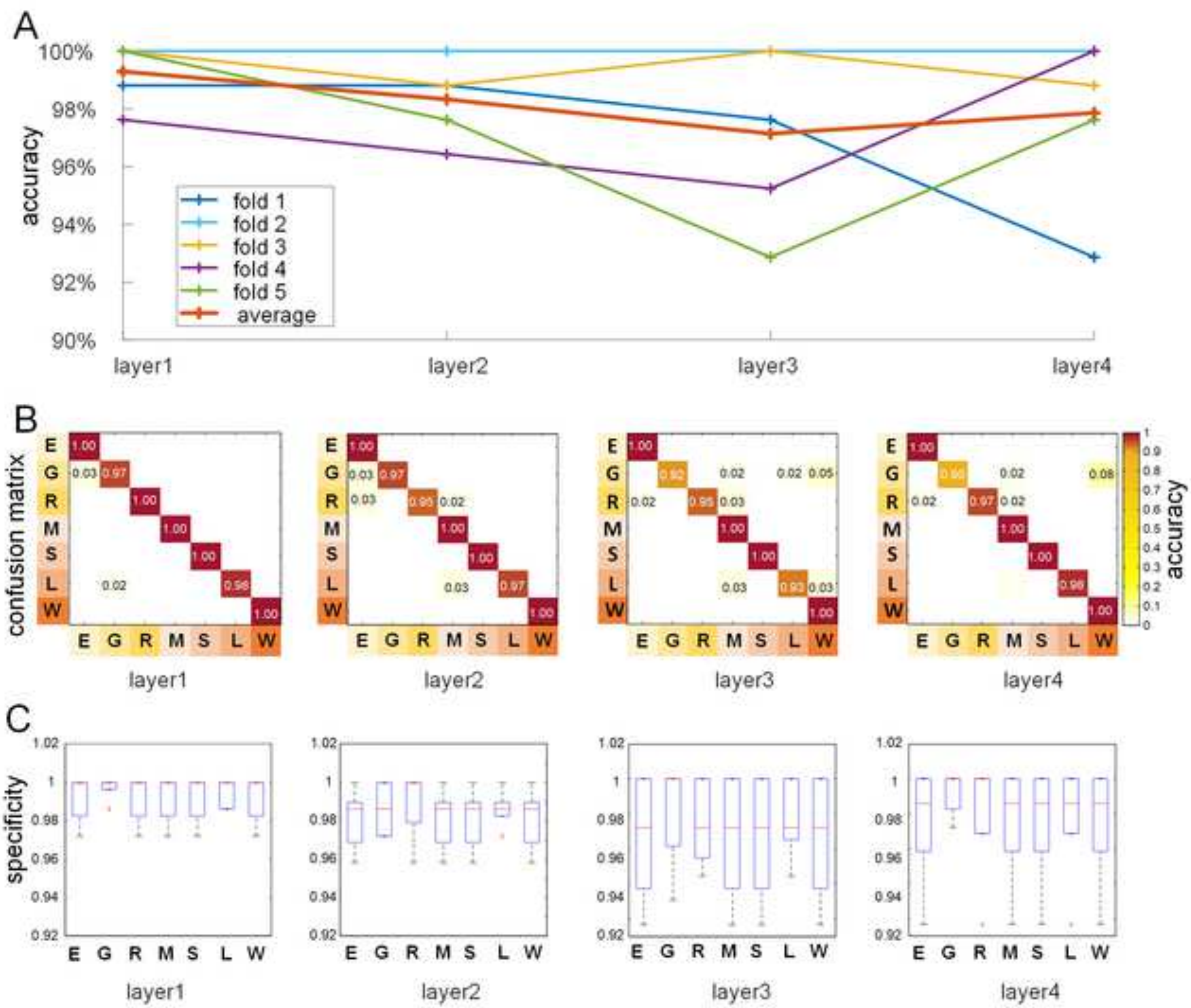
Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., & Milham, M. P. (2010). Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. Neuroimage, 49(3), 2163-2177. doi:10.1016/j.neuroimage.2009.10.080

Figure


A
Individual data preparing
4D volume for 7 tasks
Registration
MNI152 template
Signals Extraction
2D matrix for 7 tasks(S_i)
5-fold cross validation
Training set
Testing set
Training
Group-wise data
DBN
Lasso
SR
Testing
Lasso
Lasso
a² for training
a_test² for testing
Classification and ROA analysis
S
V
M
Label
ROA value computing
number of components
B
DBN
layer1
layer2
layer3
layer4
W1
W2
W3
W4
Temporal features
as layer1 D¹
as layer2 D¹
as layer3 D¹
as layer4 D¹
Lasso regression
Lasso regression
Lasso regression
Lasso regression
layer1 a¹
layer2 a¹
layer3 a¹
layer4 a¹
Sparse Representation (SR)
SR
SR
SR
layer1 D²
layer1 a²
layer2 D²
layer2 a²
layer3 D²
layer3 a²
layer4 D²
layer4 a²
FBNs
Classification performance

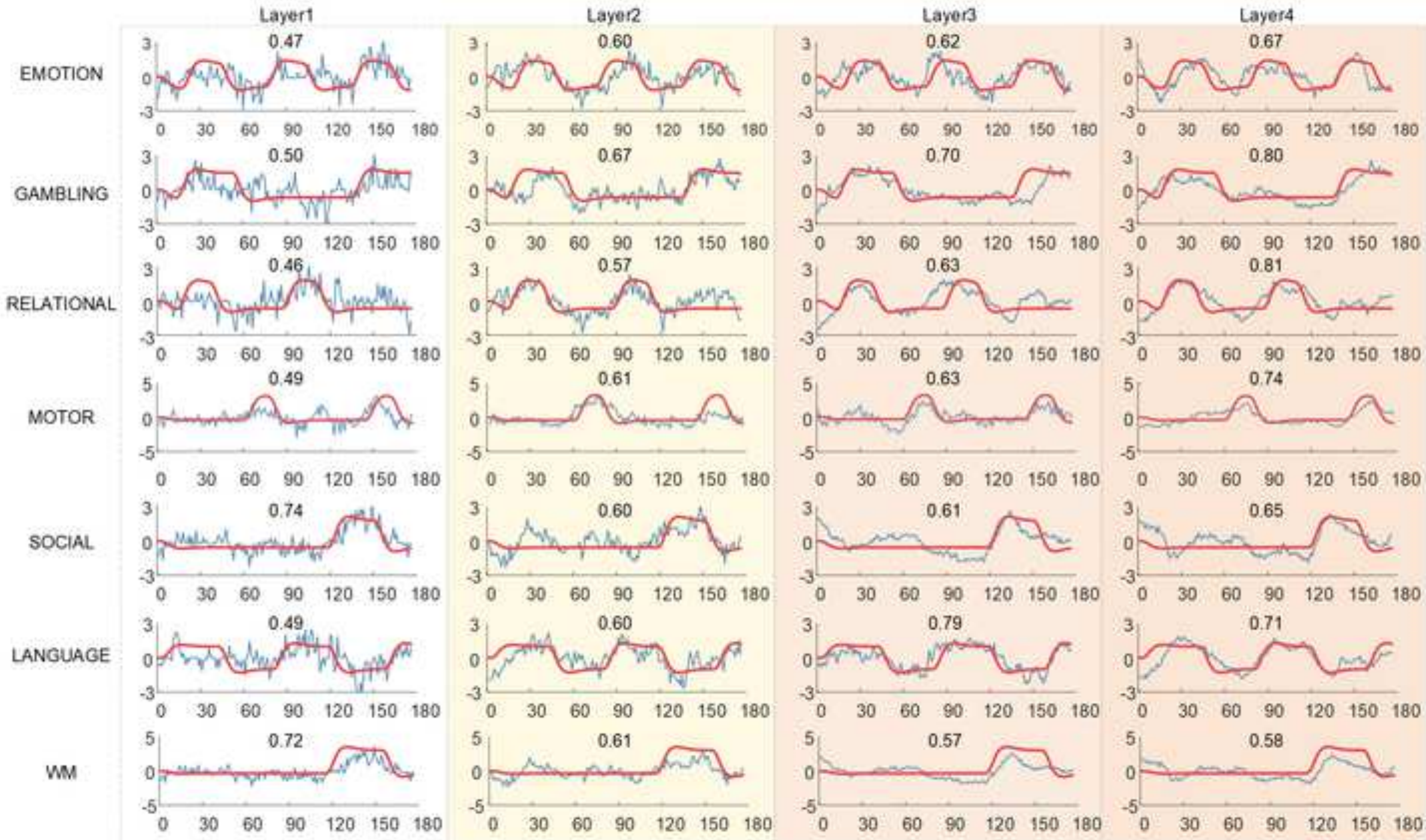
Figure

Click here to access/download;Figure;Figure 2.tif

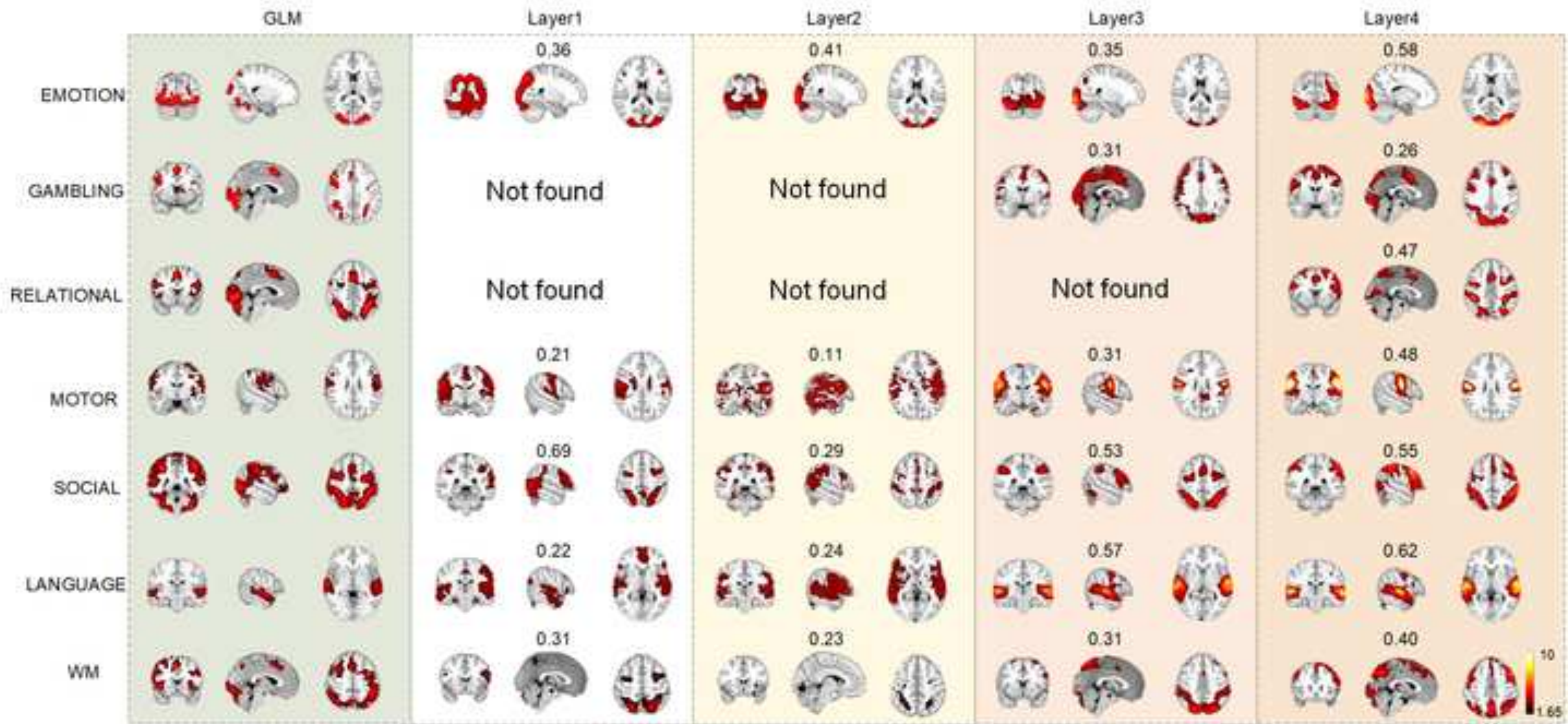


Figure

Click here to access/download;Figure;Figure 3.tif

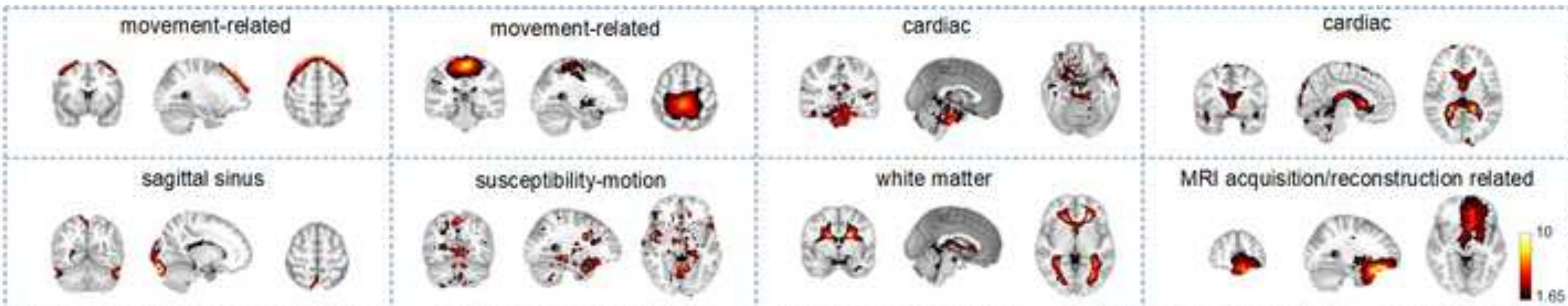


Figure

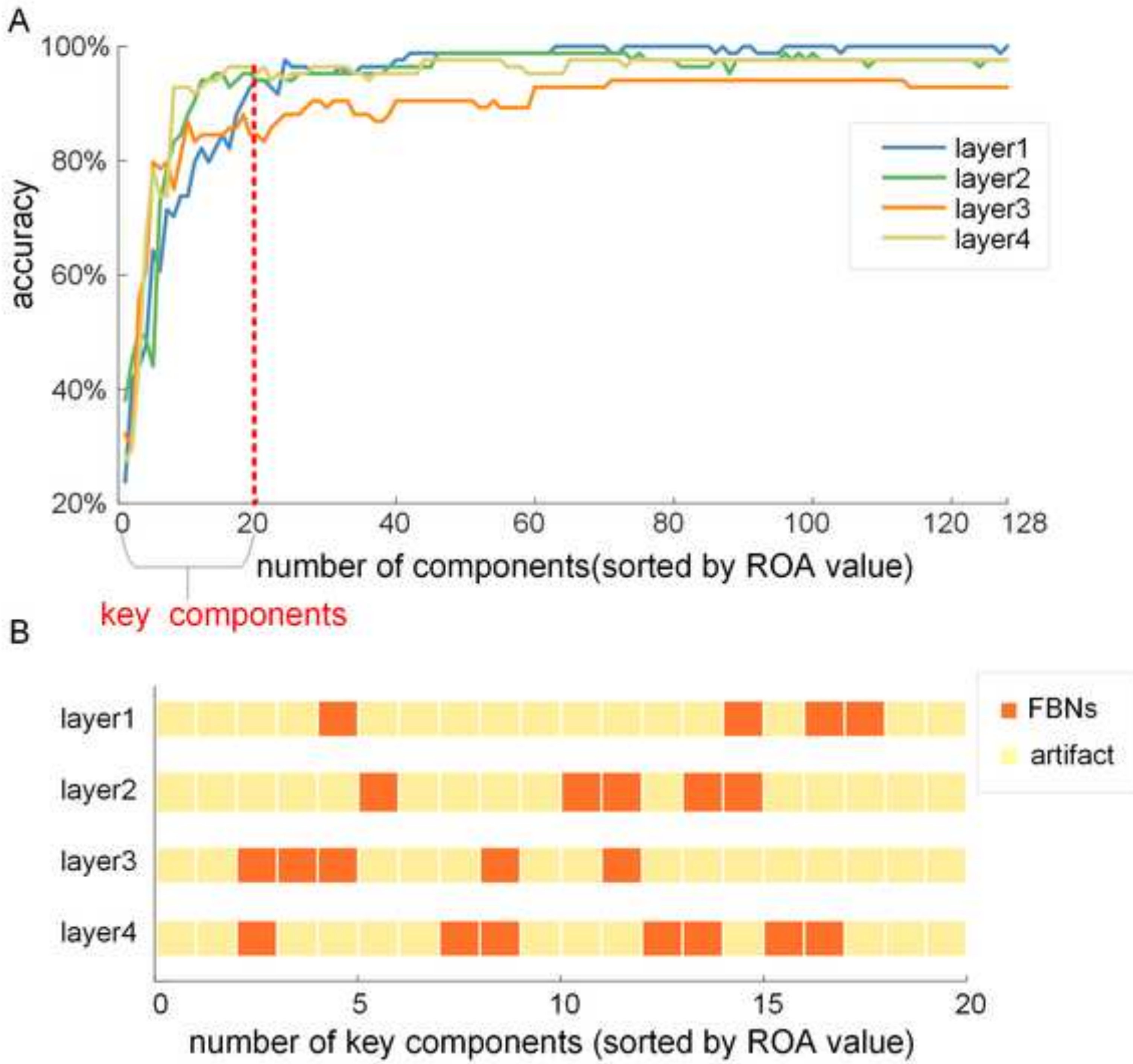

GLM
Layer1
Layer2
Layer3
Layer4
EMOTION
GAMBLING
RELATIONAL
MOTOR
SOCIAL
LANGUAGE
WM
0.36
0.41
0.35
0.58
0.31
0.26
0.47
0.21
0.11
0.31
0.48
0.69
0.29
0.53
0.55
0.22
0.24
0.57
0.62
0.31
0.23
0.31
0.40
Not found
Not found
Not found
Not found
Not found
10
1.65

Figure

Click here to access/download;Figure;Figure 5.tif



Figure


A
accuracy
100%
80%
60%
40%
20%
0
20
40
60
80
100
120
128
number of components(sorted by ROA value)
key components
layer1
layer2
layer3
layer4
B
layer1
layer2
layer3
layer4
FBNs
artifact
0
5
10
15
20
number of key components (sorted by ROA value)

Decoding different cognitive processes using task-based functional magnetic resonance imaging (tfMRI) is crucial for understanding the relationship between brain activities and cognitive states. However, existing machine learning-based feature extraction methods for decoding brain states may struggle to capture the complex and precise spatiotemporal patterns of brain activity from the highly noisy raw fMRI data. Additionally, current deep learning-based end-to-end decoding models struggle to unveil interpretable components in tfMRI signal decoding.

To address these limitations, we proposed a novel framework, the hybrid spatio-temporal deep belief network and sparse representations (DBN-SR) framework, which effectively distinguished multi-task fMRI signals with an average accuracy of 97.86%. Furthermore, it simultaneously identified multi-level temporal and spatial patterns of multiple cognitive tasks. By utilizing a novel Ratio-of-Activation metric, our framework unveiled interpretable components with greater classification capacity, offering an effective methodology for basic neuroscience and clinical research.