Brain Imaging and Behavior (2019) 13:893–904
https://doi.org/10.1007/s11682-018-9901-5

ORIGINAL RESEARCH

Encoding the local connectivity patterns of fMRI for cognitive task and state classification

Itir Onal Ertugrul[1] · Mete Ozay[2] · Fatos T. Yarman Vural[3]

Published online: 15 June 2018

Abstract

In this work, we propose a novel framework to encode the local connectivity patterns of brain, using Fisher vectors (FV), vector of locally aggregated descriptors (VLAD) and bag-of-words (BoW) methods. We first obtain local descriptors, called mesh arc descriptors (MADs) from fMRI data, by forming local meshes around anatomical regions, and estimating their relationship within a neighborhood. Then, we extract a dictionary of relationships, called *brain connectivity dictionary* by fitting a generative Gaussian mixture model (GMM) to a set of MADs, and selecting codewords at the mean of each component of the mixture. Codewords represent connectivity patterns among anatomical regions. We also encode MADs by VLAD and BoW methods using k-Means clustering. We classify cognitive tasks using the Human Connectome Project (HCP) task fMRI dataset and cognitive states using the Emotional Memory Retrieval (EMR). We train support vector machines (SVMs) using the encoded MADs. Results demonstrate that, FV encoding of MADs can be successfully employed for classification of cognitive tasks, and outperform VLAD and BoW representations. Moreover, we identify the significant Gaussians in mixture models by computing energy of their corresponding FV parts, and analyze their effect on classification accuracy. Finally, we suggest a new method to visualize the codewords of the learned brain connectivity dictionary.

Keywords fMRI · Brain decoding · Fisher vector encoding · Mesh arc descriptors

Introduction

Brain decoding methods employ brain activity records to predict information about external stimuli (Chen et al. 2014; Daliri 2014). Functional Magnetic Resonance Imaging (fMRI) is a powerful tool used for capturing neural activations observed in a wide range of cognitive tasks including object detection (Mitchell et al. 2008; Behroozi and Daliri 2014), human emotion categorization (Saarimäki et al. 2015), autobiographical memory retrieval (Rissman et al. 2016) and auditory categorization (Lee et al. 2015). Traditional approaches employ fMRI Blood Oxygenation Level Dependent (BOLD) response of voxels or anatomical regions for cognitive state classification (Haxby et al. 2001; Mitchell et al. 2004; Behroozi and Daliri 2015). Additionally, a number of feature selection methods are applied on voxel activations to select informative voxels (Daliri 2012). Yet, recent findings show that connectivity patterns observed between voxels or anatomical regions provide more information about activities performed in brain compared to the individual voxel BOLD responses. In addition, connectivity between BOLD responses, which are frequently used in resting-state fMRI analysis (Khazaee et al. 2016; Cai et al. 2017), has been shown to provide better classification performance compared to traditional approaches (Richiardi et al. 2011; Shirer et al. 2011). Brain connectivity is also represented by a set of local meshes (Onal et al. 2015a, b), where relationships among multiple voxels are estimated within a predefined neighborhood. Estimated relationships, called mesh arc descriptors (Ozay et al. 2012), are reported to give the best performance

✉ Itir Onal Ertugrul
iertugru@andrew.cmu.edu

Mete Ozay
mozay@vision.is.tohoku.ac.jp

Fatos T. Yarman Vural
vural@ceng.metu.edu.tr

[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

[2] Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan

[3] Department of Computer Engineering, Middle East Technical University, Ankara, Turkey

in cognitive state classification, compared to voxel BOLD responses and their pairwise relationships.

Encoding approaches are widely used in pattern recognition literature to improve representation capacity of local descriptors. In a popular encoding approach, called bag-of-words (BoW), first local descriptors are clustered using a k-means clustering method. Then, cluster centroids are defined as codewords to form a dictionary. Codewords are utilized as *textual words* in natural language processing, and *visual words* in image processing. Each descriptor is assigned to the closest codeword, and a sample is represented by a histogram of codewords, to which its local descriptors belong. BoW approach has been used to detect diseases (Solmaz et al. 2012) and cognitive states (Sucu et al. 2016) using fMRI data, and to classify EEG time series data (Wang et al. 2013; Meriño et al. 2013).

In another encoding approach, called vector of locally aggregated descriptors (VLAD), first, codewords are computed similar to BoW. However, VLAD aims to accumulate the difference between codewords and local descriptors assigned to the codewords (Jégou et al. 2010). VLAD has been used in various applications including image (Delhumeau et al. 2013) and video (Abbas et al. 2015) processing, yet it has not been used to encode local descriptors obtained from fMRI data.

Fisher vector (FV) encoding methods are employed for statistical data analysis by making use of the benefits of generative and discriminative models (Carvajal et al. 2016; Sánchez et al. 2013). FV encoding is considered as an extension of BoW such that, rather than encoding the relative frequency of the descriptors, it encodes the information on distribution of the descriptors (Perronnin et al. 2010). Fisher kernels have been used to compute FVs utilizing a mechanism that incorporates generative probability models into discriminative classifiers in various applications, including classification of protein domains (Jaakkola et al. 1999), action and event recognition (Oneata et al. 2013; Sekma et al. 2015), image classification (Liu et al. 2014; Simonyan et al. 2013; Sánchez and Redolfi 2015) and 3D object retrieval (Savelonas et al. 2016). FVs have been applied to model effective connectivity of networks using MRI and PET data (Zhou et al. 2016). Yet, FVs have not been used to encode the connectivity patterns of fMRI data, which is very crucial to analyze the behavior of brain during cognitive tasks.

In this study, our aim is to show that generating a *brain connectivity dictionary* and encoding brain connectivity patterns using generative embedding methods provide more discriminative information compared to modeling connectivity patterns obtained for individual samples. We show the significance of generative embedding of connectivity patterns on two different brain decoding problems, namely, decoding different tasks and decoding cognitive states within a task.

The major contribution of this study is to suggest a novel framework which models the connectivity patterns among the anatomic regions by encoding a set of local descriptors using the fMRI data. Inspired by the concept of visual words in computer vision and pattern recognition, our motivation is to generate a *brain connectivity dictionary* that represents relational patterns formed among anatomical regions in order to describe the cognitive tasks and states. The proposed framework enables us to examine various encoding methods such as FVs, VLAD and BoW for cognitive task and state classification. To our knowledge, our work provides the first framework which employs encoding methods for connectivity analysis of fMRI data. We train linear SVM classifiers using the encoded MADs which are obtained from BoW, VLAD and FV encoding methods. We observe that encoding MADs using FV encoding provides the best performance, and improves the performance of raw MADs in decoding different tasks of Human Connectome Project (HCP) task-fMRI dataset and in decoding emotional vs. neutral states of Emotional Memory Retrieval dataset.

Another contribution of this work is the exploration of the relationship between energy of FV columns obtained from a particular Gaussian component and the classification accuracy. Our results reflect that Gaussian mixture components whose FV columns have higher energy are more discriminative compared to the ones with lower energy. Finally, we suggest a visualization method to depict and analyze codewords of the brain connectivity dictionary on a human brain template. We first sort the Gaussian mixture components based on the energy they provide for FV encoding. Then, we plot their codewords on brain, and sort them with respect to their discriminative power.

Data representation

Let a time series, obtained at each voxel v, be represented by a function $X_v(t)$ of time t. If we work at voxel granularity level, then nodes are determined by voxels, and the corresponding node time series is defined by $X_i(t) \triangleq X_v(t)$ for node i. On the other hand, we obtain a representative region time series $X_r(t)$ for each anatomical region r by spatially averaging the time series of voxels residing in that region using

$$X_r(t) = \frac{1}{V_r} \sum_{v \in r} X_v(t), \tag{1}$$

where V_r denotes the number of voxels residing in region r. If we work in anatomical region granularity, then nodes are determined by regions, and the corresponding node time series is defined by $X_i(t) \triangleq X_r(t)$. In the

experimental analyses, we use two datasets called the Human Connectome Project task-fMRI dataset and the Emotional Memory Retrieval (EMR) dataset, which are explained below.

Human connectome project task-fMRI dataset

We first use the benchmark task fMRI dataset of Human Connectome Project (HCP) (Barch et al. 2013). We employ task fMRI data of 100 healthy subjects collected for seven cognitive tasks, namely, Emotion (Emo), Gambling (Gam), Language Processing (Lan), Motor (Mot), Relational Processing (Rel), Social Processing (Soc) and Working Memory (WM). Summary of demographics information of subjects can be found in Table 1. Number of scans and their duration vary for each task, yet they have equal duration for all participants (see Table 2).

In this dataset, nodes correspond to regions and for each node $i = 1, 2, \ldots, N$, a time series is represented by $X_i(t) = X_r(t)$. We use $N = 90$ regions of Automated Anatomical Labeling (AAL) brain atlas (Tzourio-Mazoyer et al. 2002) except the ones that cover Cerebellum and Vermis.

Emotional memory retrieval (EMR) dataset

In this experiment, the stimuli consist of neutral (Kitchen utensils and Furniture) and emotional (Fear and Disgust) categories of images. Within each trial, participants were presented with 5 images from the same category. Then, they solved three simple math problems. After that, a 2 seconds retrieval period was started. In each period, participants were presented with a test image from the same category, and decided whether the image was a member of the current study list or not. For more detailed experimental setting, please refer to Mızrak et al. (2017). This experiment consisted of six 19-min runs, and each run contains 35 trials. In total, 210 trials were obtained for a single participant. We collected data from 13 participants whose demographics information is provided in Table 1. Participants had normal or corrected to normal vision and were screened for medical conditions that could contradict with MRI protocols. We employ 6 measurements (6 seconds of encoding and 6 seconds of following rest) of each trial for classification.

Table 1 Demographics information of subjects in HCP and EMR datasets

	Age Range	Gender	Race	% of Healthy
HCP	22 - 35	54 Female, 46 Male	N/A	100
EMR.	22 - 27	8 Female, 5 Male	White	100

Table 2 Number of scans obtained per session, and its duration (min:sec)

	Emo	Gam	Lan	Mot	Rel	Soc	WM
Scans	176	253	316	284	232	274	405
Dur.	2:16	3:12	3:57	3:34	2:56	3:27	5:01

Notice that, we only employ the data obtained during the encoding phase.

In this dataset, nodes correspond to voxels, and the corresponding time series is denoted by $X_i(t) \triangleq X_v(t)$ for each node $i = 1, 2, \ldots, N$. The active voxels are scattered all over the brain regions in the EMR dataset. Since voxels provide more discriminative information compared to regions, we work in finer-granularity with this dataset, and nodes correspond to voxels as utilized in the work of Onal et al. (2017). The most discriminative voxels are selected by a t-test among all the voxels of this dataset. In the test, we first split samples into two groups according to their class labels. Then, we compute p-values for each voxel, which reflect how the groups are well-separated according to activations obtained from each voxel. We select $N = 100$ voxels, since almost 100 of them have $p < 0.001$. We perform voxel selection repetitively for each training split of cross validation.

Encoding of mesh arc descriptors

In this section, we introduce our framework to encode MADs. First, we estimate MADs to represent the local connectivity patterns of cognitive tasks. Next, we change the MAD space using PCA and ICA. Then, we encode MADs using three methods by employing their distance to the centers of GMMs, namely, Fisher vectors (FV), vector of locally aggregated descriptors (VLAD) and bag-of-words (BoW). The encoded MADs are used to train an SVM classifier for cognitive task classification. Figure 1 shows an overview of our encoding framework.

Estimation of mesh arc descriptors (MADs)

Suppose that, we record the BOLD response at a node i given to a stimulus within a window w, in order to measure the brain activation for a predefined cognitive state with label c. We denote the intensity values of the BOLD response measured at a node i within a window w by the vector $\boldsymbol{x}_{w,i} = [x_{w,i}(t)]_{t=1}^{D_w}$, where we record D_w measurements for each window. Notice that, the number of the measurements D_w recorded within the window w depends on the class of the trial for HCP dataset. For instance, we record $D_w \in (176, 253, 316, 284, 232, 274, 405)$ number

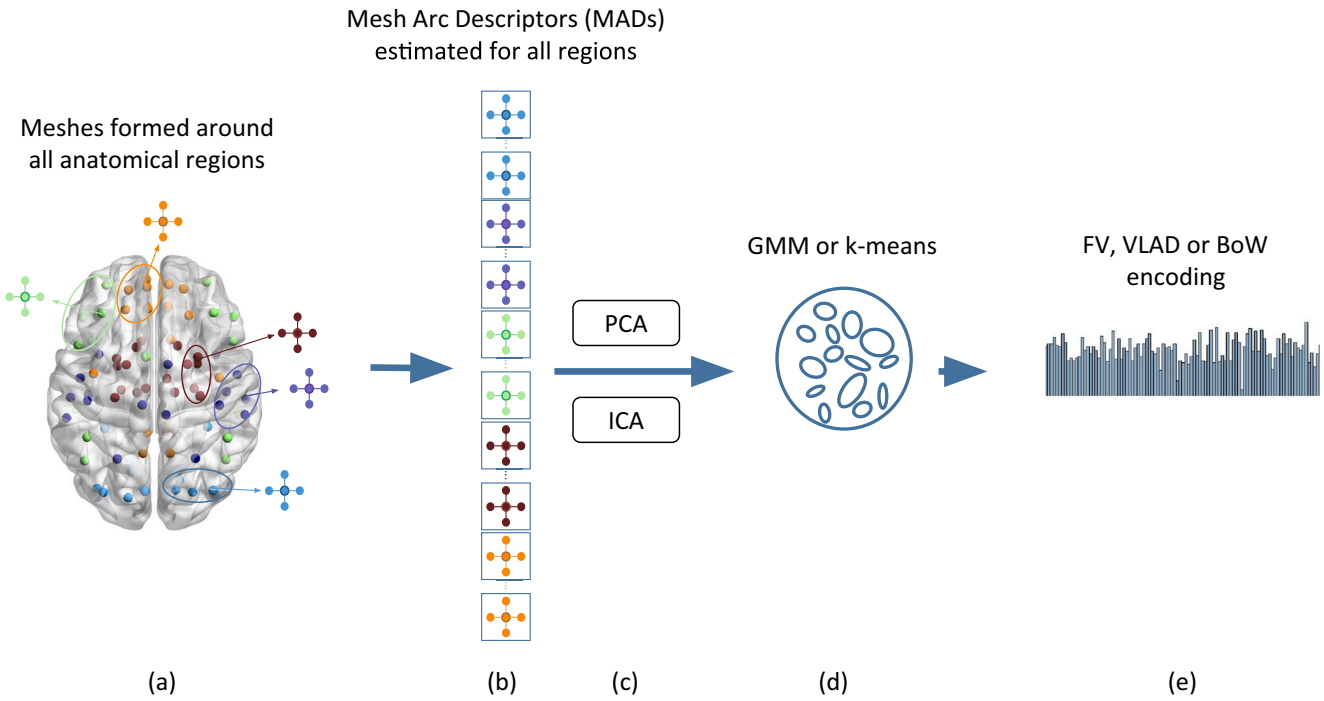

Fig. 1 An overview of our proposed framework which is employed to generate a brain connectivity dictionary. For each cognitive task, we first form local meshes around all anatomical regions using their functionally nearest neighbors (in the figure, we represent only five representative meshes). Next, we estimate edge weights of meshes (MAD) for all tasks of all subjects. We apply PCA or ICA to obtain decorrelated descriptors. Then, we employ GMMs and k-means clustering methods on the descriptors using data obtained from all tasks. Finally, we obtain FV, VLAD and BoW encoding for each task

of measurements for $c = 1, 2, \ldots, 7$ as denoted in Table 2. On the contrary, for EMR dataset, window size equals to $D_w = 6$ for all of the 210 trials independent of the class label $c = 1, 2$.

We form meshes around nodes for each window by connecting each node to its functionally nearest neighboring nodes. Functionally p-nearest neighbors of a node are defined by the nodes whose time series have the p-largest Pearson correlation coefficient associated to time series of the given node during a cognitive process. For a window, we denote the estimated weight of an arc between seed node of a mesh i and a neighboring voxel j by $a_{w,i,j}$.

We represent a mesh formed among the BOLD response of a seed node i of the mesh, and the BOLD responses of its p-nearest neighbors $\eta_p[i]$ for a window w by

$$\hat{\boldsymbol{x}}_{w,i} = \sum_{j \in \eta_p[i]} a_{w,i,j}\, \boldsymbol{x}_{w,j}, \tag{2}$$

where $\hat{\boldsymbol{x}}_{w,i}$ denotes the estimated BOLD response vector of node i for window w. We estimate the arc weights of a mesh formed around a node i for a window w by minimizing the regularized linear model error as follows:

$$\min_{a_{w,i,j}} \left(||\hat{\boldsymbol{x}}_{w,i} - \boldsymbol{x}_{w,i}||_2^2 + \lambda ||a_{w,i,j}||_2^2 \right), \tag{3}$$

where $||\cdot||_2^2$ is the squared ℓ_2 norm and $\lambda \in \mathbb{R}$ denotes regularization parameter.

For each $\boldsymbol{a}_{w,i}$, we allocate a MAD vector of size $1 \times N$, where N is the number of nodes, and fill the corresponding entries with the mesh arc descriptors $a_{w,i,j}$. In other words, $\boldsymbol{a}_{w,i}(j) = a_{w,i,j}$ if $j \in \eta_p[i]$ and $\boldsymbol{a}_{w,i}(j) = 0$ if $j \notin \eta_p[i]$.

For simplicity, while encoding MADs, we remove window w subscript of $\boldsymbol{a}_{w,i}$ in Eq. 2, and denote a MAD vector by $\boldsymbol{a}_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,N}]$ of size $1 \times N$, for each trial. Then, all the MADs are concatenated under a vector of size $1 \times N^2$, such that $\mathbf{a} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$, for all trials of all subjects.

Encoding methods

In this subsection, we explain how MADs are encoded for fMRI data analysis.

Fisher Vector (FV) encoding

Given a set of vectors, Fisher Vector (FV) encoding method is used to encode deviation of distribution of the vectors from a dictionary, which is typically described by a diagonal GMM (Carvajal et al. 2016). In the proposed framework, our vectors are MADs obtained from fMRI

data. In order to generate the *brain connectivity dictionary*, we fit a Gaussian mixture model (GMM) to the set of MADs, obtained from all anatomic regions, and from all subjects, during all cognitive tasks. The mean vector of a Gaussian component corresponds to a codeword of the *brain connectivity dictionary*. In order to satisfy the assumption of diagonal covariance matrix of the GMM and obtain linearly uncorrelated features, we apply PCA or ICA to MADs. For simplicity, we use the same notation for MADs at the output of PCA, such that $\mathbf{a}_i$ represents the projection of a MAD in a D-dimensional space.

Let $A = \{\mathbf{a}_i \in \mathbb{R}^D; i = 1, 2, \ldots, N\}$ denote a set of D-dimensional MADs which are obtained from a single task of a single subject, and are sampled from the set of all MADs, $\mathbb{A}$. Also, let u_k be the k^{th} component of the GMM, which models the generative process for elements belonging to $\mathbb{A}$. We denote the parameter set of all u_k's by $\theta = \{\omega_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k; k = 1, 2, \ldots, K\}$, where ω_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mixture weight, mean vector and covariance matrix of the k^{th} Gaussian component, respectively. Note that, K denotes the number of Gaussian components in the mixture. We compute the mixture u_θ by $u_\theta(a) = \sum_{k=1}^{K} \omega_k u_k(a)$, where $u_k(a)$ denotes the k^{th} Gaussian in the mixture, and is computed by

$$u_k(a) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(a-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(a-\boldsymbol{\mu}_k)\right). \tag{4}$$

We estimate parameters of the GMM on a training set of MADs using an Expectation Maximization (EM) algorithm. We assume that diagonal covariance matrices are identified by $\boldsymbol{\sigma}_k^2$ which denotes a vector of the diagonal entries of $\boldsymbol{\Sigma}_k^2$. Then, we compute the derivatives of log-likelihood of the GMM with respect to the parameters using $\boldsymbol{G}_\theta^A = N^{-1} \sum_{i=1}^{N} \nabla_\theta \log u_\theta(\mathbf{a}_i)$. The derivatives are only taken with respect to Gaussian means and variances, since the gradients computed with respect to the weight parameters, ω_k, provide *less information* (Perronnin et al. 2010). Therefore, we obtain a representation that captures the average first and second order differences between MADs and each of the GMM centers. D-dimensional gradients are computed by

$$\mathcal{G}_{\boldsymbol{\mu}_k}^A = \frac{1}{N\sqrt{\omega_k}} \sum_{i=1}^{N} \gamma_i(k) \left(\frac{\mathbf{a}_i - \boldsymbol{\mu}_k}{\boldsymbol{\sigma}_k}\right) \tag{5}$$

and

$$\mathcal{G}_{\boldsymbol{\sigma}_k}^A = \frac{1}{N\sqrt{2\omega_k}} \sum_{i=1}^{N} \gamma_i(k) \left[\frac{(\mathbf{a}_i - \boldsymbol{\mu}_k)^2}{\boldsymbol{\sigma}_k^2} - 1\right], \tag{6}$$

where $\gamma_i(k)$ represents the soft assignment of the MAD vector $\mathbf{a}_i$ to the k^{th} Gaussian, and is defined by $\gamma_i(k) = \omega_k u_k(\mathbf{a}_i) \left(\sum_{l=1}^{K} \omega_l u_l(\mathbf{a}_i)\right)^{-1}$ (Perronnin and Dance 2007).

A Fisher vector $\boldsymbol{G}_\theta^A$ is obtained by concatenating the gradients under a vector $\boldsymbol{G}_\theta^A = \left[\mathcal{G}_{\boldsymbol{\mu}_1}^A, \mathcal{G}_{\sigma_1}^A, \mathcal{G}_{\boldsymbol{\mu}_2}^A, \mathcal{G}_{\boldsymbol{\sigma}_2}^A, \ldots, \mathcal{G}_{\boldsymbol{\mu}_K}^A, \mathcal{G}_{\boldsymbol{\sigma}_K}^A\right]$ of size $2KD$. We obtain FVs for both training and test data. In order to obtain better accuracy, we further apply l_2 normalization and square-root transformation on FVs as suggested in Sánchez et al. (2013).

The code can be found at https://github.com/itironal/FisherVector-mesh.

Vector of locally aggregated descriptors (VLAD)

Vector of Locally Aggregated Descriptors (VLAD) encoding method is used to encode a set of descriptors into a dictionary, which is computed by a k-means clustering method. In the proposed framework, we first perform k-means clustering of MADs on the training data. Hence, the cluster centroids $\boldsymbol{\mu}_k$, $k = 1, 2, \ldots, K$, correspond to our codewords of *brain connectivity dictionary*. Then, we associate each MAD $\mathbf{a}_i$ to its nearest codeword $NN(\mathbf{a}_i)$ in the dictionary, such that $NN(\mathbf{a}_i) = \underset{\boldsymbol{\mu}_k}{\operatorname{argmin}} \|\mathbf{a}_i - \boldsymbol{\mu}_k\|_2$, where $\|\cdot\|_2$ is the Euclidean norm.

Recall from the previous subsection that, $A = \{\mathbf{a}_i \in \mathbb{R}^D; i = 1, 2, \ldots, N\}$ represents the set of D-dimensional MADs belonging to a single task of a single subject, each of which is sampled from $\mathbb{A}$ that contains all MADs belonging to all tasks of all subjects. For each codeword $\boldsymbol{\mu}_k$, we compute the sum of the differences, $(\mathbf{a}_i - \boldsymbol{\mu}_k)$, of the descriptors $\mathbf{a}_i$, which are assigned to the k^{th} cluster by

$$\mathfrak{v}_k^A = \sum_{\mathbf{a}_i : NN(\mathbf{a}_i) = \boldsymbol{\mu}_k} (\mathbf{a}_i - \boldsymbol{\mu}_k). \tag{7}$$

We concatenate the computed D-dimensional vectors $\mathfrak{v}_k^A$ for all clusters, and obtain a KD dimensional VLAD encoding $\boldsymbol{\mathcal{V}}^A = [\mathfrak{v}_1^A, \mathfrak{v}_2^A, \ldots, \mathfrak{v}_K^A]$. Note that, we obtain a VLAD encoding for each A, which is computed using training and test data.

Bag-of-words (BoW)

We also encode MADs using a Bag-of-Words (BoW) method, where words correspond to a set of selected MADs. We first cluster MADs obtained from training data using a k-means clustering method. Then, for each A, we compute the number of MADs, b_k^A, belonging to the k^{th} cluster. Finally, we represent each sample by a K-dimensional vector such that $\boldsymbol{B}^A = [b_1^A, b_2^A, \ldots, b_K^A]$.

Experimental results

We perform two sets of experiments. First, we classify cognitive tasks of the HCP dataset and cognitive states of the EMR dataset using the raw and encoded MADs to see the effect of encoding on MADs, and compare the performance of various encoding methods. In addition, we compare our proposed methods with the state-of-the-art fMRI data representation methods for classification of cognitive states and tasks. Finally, we visualize and qualitatively analyze codewords of *brain connectivity dictionaries* computed using the proposed methods.

Classification results

In order to perform classification, we first computed FV ($\boldsymbol{G}_{\theta}^{A}$), VLAD ($\boldsymbol{\mathcal{V}}^{A}$) and BoW ($\boldsymbol{B}^{A}$) encoding of local MADs for all tasks and all participants. Then, we used the encoded MAD vectors to train and test linear SVM classifiers. We measured the classification performance of the HCP dataset using a 10-fold cross validation (CV) scheme by randomly splitting the data into 10 subsets according to a uniform distribution, training each model on 9 splits, and testing the learned model on the remaining split. Note that, training and test splits contain data associated to different participants for each fold. For the EMR dataset, we performed 13-fold cross-validation, where each split contains data collected from a participant. We trained each model on 12 splits, and tested the learned model on the 13^{th} split. In the proposed framework, we used only training data for implementation of all methods that were employed in learning phase, such as computation of PCA and ICA using MADs, estimation of GMMs, k-means clustering and training SVM classifiers. We examined the proposed methods for various number of neighbors $p \in P = \{10, 20, 30, 40\}$, in each mesh to observe the effect of degree of locality of the meshes. Considering the fact that MADs are N-dimensional vectors, employment of smaller p values provides increased locality and sparser MAD vectors. Before employment of FV encoding, we applied PCA or ICA to MADs for decorrelation purpose, and did not change their dimension ($D = N$). Moreover, we selected the number of Gaussians in GMM and the number of clusters in k-means using $k \in \{20, 40, 60, 80, 100, 120\}$. We share the results of k, whose average performance over all folds is optimal. Accuracy values given in Table 3 show the performances obtained with optimal values of k found using cross validation.

Results given in Table 3 show that, FV encoding of MADs computed after employment of PCA gives the best performance to classify cognitive tasks compared to VLAD and BoW encoding. Unlike PCA, applying ICA on MADs before encoding with FV, VLAD and BoW decreases the performance of MADs. We observe that VLAD encoding of MADs does not increase the discrimination power of MADs. Although accuracy values of VLAD are comparable with that of MAD, BoW representations of MADs yield substantial decrease in accuracy.

We observed that optimal k value of k-Means or GMM varies for each p and each of the methods meaning that the optimal k is not constant for all methods. In order to interpret the relationship between k and classification performances, we analyze the standard deviations of performances obtained with all values of k for the given methods. We plot the standard deviations of performances obtained for $k \in K$ for each method in Fig. 2. We observe that standard deviations of performances obtained with FV based methods are lower compared to others meaning that FV-based methods are more robust to k.

We also give the running time of each step of our framework for $p = 40$ on the HCP dataset in Table 4.

Table 3 Classification performance (%) computed for different encoding methods obtained using the HCP and the EMR datasets

	HCP				EMR			
p	10	20	30	40	10	20	30	40
MAD	88.14	91.14	93.00	93.86	71.17	71.10	70.15	71.10
FV-PCA	**91.71**	**94.71**	**95.29**	**95.71**	**75.97**	**76.08**	**75.05**	**75.53**
FV-ICA	87.29	85.29	87.71	87.76	73.88	73.41	71.61	71.79
FV-None	86.00	89.29	91.71	92.14	73.81	72.27	71.76	71.86
VLAD-PCA	89.86	91.43	92.57	93.43	74.91	74.47	74.62	73.70
VLAD-ICA	84.29	88.29	89.86	90.00	71.10	71.14	71.90	71.25
VLAD-None	86.71	89.00	90.43	91.14	72.09	71.76	72.09	72.34
BoW-PCA	62.86	68.57	67.29	68.71	64.76	64.91	63.88	63.55
BoW-ICA	41.57	42.14	46.14	45.00	59.67	57.00	58.42	57.25
BoW-None	54.57	57.86	60.43	60.29	60.62	58.63	61.61	61.72

Results in bold represent the highest accuracy among all methods

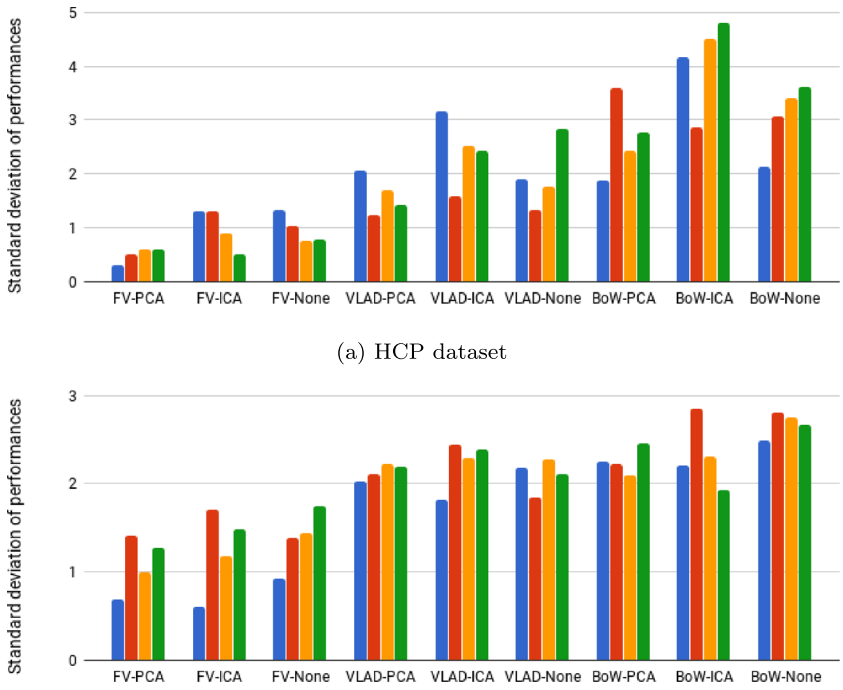

Fig. 2 Standard deviation of classification performances obtained with values of $k \in K$ for each method

(a) HCP dataset

(b) EMR dataset

It is observed that computation of ICA both takes a large amount of time and decreases the encoding performance. On the other hand, computation of PCA takes much shorter time, while contributing a lot to the classification accuracy. Moreover, we observe that computation times of FV, VLAD and BoW are less compared to those of GMM and k-means.

In order to compare the representation power of our framework, we compared our results with the state-of-the-art methods in Table 5. Representations computed using these methods include average region BOLD responses, Pearson and partial correlations, mutual information and Granger causality between BOLD responses. We further apply PCA and ICA on average BOLD responses to reduce dimension of our feature space. When the average region BOLD responses are employed for classification, we obtain 14.29% accuracy for the 7-class HCP dataset and 51.24% accuracy for the 2-class EMR dataset. These results are almost equal to Zero Rule (Zero-R) accuracy values in which the classifier predicts the majority class for the HCP (14.29%) and the EMR (50%) datasets, respectively.

Table 4 Running time of each step of the framework measured in seconds

MAD	PCA / ICA	GMM / k-means	FV / VLAD / BoW
105.6	8.4 / 373.6	211.3 / 62.8	14.3 / 7.7 / 11.9

Since average BOLD signals do not carry significant information, we also obtain bad results when we apply PCA and ICA on them. On the other hand, when pairwise correlations between all pairs of region BOLD responses are used as features for the HCP dataset, we obtain a classification accuracy of 77.42% using Pearson correlation, and 67.86% using partial correlation. Since the EMR dataset is event-related, partial correlation cannot be computed for this dataset, yet using Pearson correlation gives 66.19% accuracy. Employing Granger causality gives Zero-R results for both datasets while employing mutual information gives 27.85% on the HCP dataset, and 64.32% on the EMR

Table 5 Comparison of classification accuracy (%) with the state-of-the-art methods

	HCP	EMR
BOLD	14.29	51.24
BOLD + ICA	14.29	52.27
BOLD + PCA	14.14	51.32
Pearson Corr.	77.42	66.19
Partial Corr.	67.86	–
Granger Causality	19.61	50.29
Mutual Information	27.85	64.32
MAD	93.86	71.17
FV+MAD+PCA	**95.71**	**76.08**

Results in bold represent the highest accuracy among all methods

Table 6 Results obtained by employment of McNemar's test with Yates' correction to compare significance of using Fisher vector encoding of MADs

Dataset	p	N_{ss}	N_{ff}	N_{sf}	N_{fs}	χ^2	p-value
HCP	10	586	27	56	31	6.621	0.01008
	20	626	25	37	12	11.76	0.00061
	30	638	20	29	13	5.36	0.02064
	40	647	20	23	10	4.36	0.03671
EMR	10	1724	437	350	219	29.70	000001
	20	1740	452	337	201	33.88	0.00001
	30	1712	478	337	203	32.76	0.00001
	40	1733	461	328	208	26.42	0.00001

dataset. In Table 5, we observe a substantial performance increase ($10\% - 15\%$) using MADs. Note that, we share the results of MADs obtained using $p = 40$ for the HCP dataset and $p = 10$ for the EMR dataset, which are shown to give the best results in Table 3. Finally, we obtain ($2\% - 5\%$) performance increase by encoding MADs using FV. Here, we share the results obtained using ($p = 40, k = 40$) and ($p = 20, k = 100$) for HCP and EMR datasets, respectively.

In order to quantitatively analyze superiority of the proposed FV encoding of MADs over MADs, we further employ McNemar's test with Yates' correction, which is a first order test of the statistical significance of an observed difference in recognition performance (Demšar 2006). It uses the times of success/failure of trials of the compared algorithms. Let N_{sf} and N_{fs} denote the times at which the first algorithm succeeds but the second fails and the second algorithm succeeds but the first fails, respectively. Then, we compute χ^2 as follows:

$$\chi^2 = \frac{(|N_{sf} - N_{fs}| - 1)^2}{\sqrt{N_{sf} + N_{fs}}}. \tag{8}$$

After χ^2 value is computed, the corresponding p-value is computed using the χ^2 distribution table. If the p-value is below the significance level 0.05, then the performance difference between two compared methods is considered to be statistically significant. Results given in Table 6 reflect that encoding MADs with Fisher vectors significantly improves the accuracy of MADs for all neighborhood sizes (p) of both datasets.

Next, we examine the success of encoding for each cognitive task. We observe in Fig. 3 that the accuracy value obtained for Gambling task is lower compared to the other tasks performed using FV-MAD and VLAD-MAD. However, accuracy values obtained from BoW-MAD for all classes are low, meaning that BoW is not successful to encode MADs. When we compare the results shown in

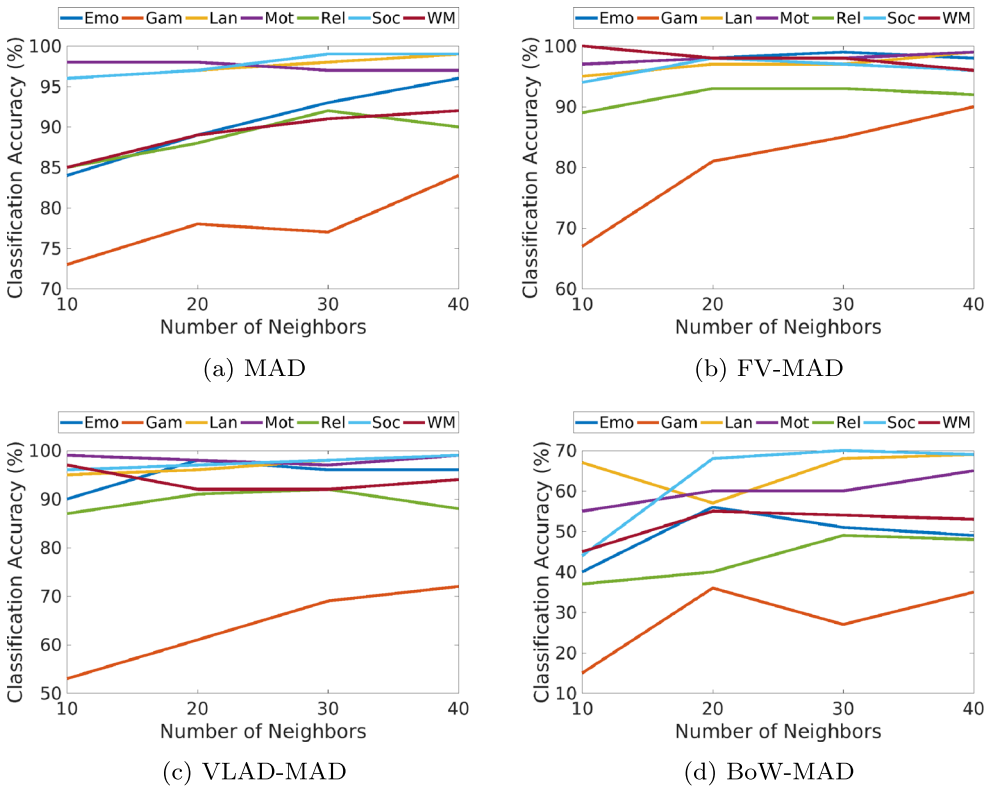

Fig. 3 Classification accuracy (%) measured for each task of the HCP dataset

Fig. 3a, b and c, we observe that we can boost the accuracy for the Emotion, Relational processing and WM tasks compared to concatenated MADs by encoding MADs with FV and VLAD. Yet, the accuracy is decreased for the Gambling task, especially for the lower values of p.

Analysis of codewords

In this subsection, we explore the contribution of a particular codeword (a component of a GMM) to classify the cognitive tasks by measuring the energy (Euclidean norm) of the corresponding columns of FVs. Recall that, the columns of the k^{th} Gaussian used in a FV encoding are denoted by $[\mathcal{G}^A_{\boldsymbol{\mu}_K}, \mathcal{G}^A_{\boldsymbol{\sigma}_K}]$ for a set A of MADs. If we denote the corresponding columns of the k^{th} Gaussian in the FV encoding using all training data by $\mathcal{G}_k = [\mathcal{G}_{\boldsymbol{\mu}_K}, \mathcal{G}_{\boldsymbol{\sigma}_K}]$, then the energy is measured by the Euclidean norm $\|\mathcal{G}_k\|_2$. We denote a Gaussian whose FV columns have the lowest energy by GLE, and the one whose FV columns have the highest energy by GHE.

Next, we visualize N-dimensional codewords obtained from MADs without using PCA or ICA. Each codeword corresponds to a mesh pattern formed around a seed voxel. Large values of codeword elements represent that a MAD formed between the region corresponding to the codeword element and the seed region has large value. Similarly, small values of codeword elements indicate small values of MADs between the seed region and the corresponding region. As an example, we plot codewords for $p = 40$ and $k = 20$, in Fig. 4. Notice from Fig. 4 that, codewords represent different mesh patterns. For example, Codeword1 represents a mesh pattern in which MADs formed between a seed region and regions residing in the occipital lobe have large values while MADs formed between a seed region and the other regions have smaller values. Recall that, each GMM component obtained without using PCA corresponds to $2N$ columns of FVs. In Fig. 4, we provide visualization of the Gaussians in an order from GHE to GLE by plotting the codewords of the corresponding Gaussians. The results show that codeword of GHE has large values mainly in

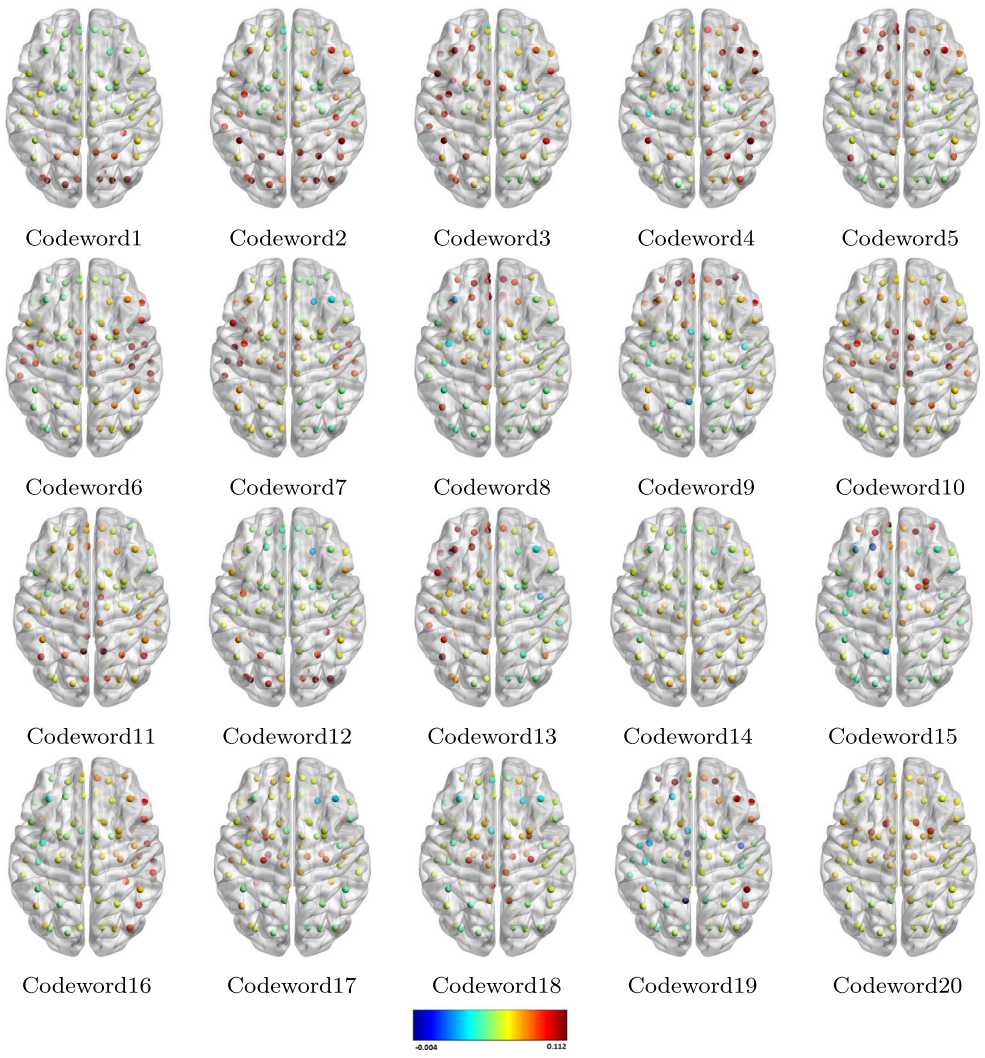

Fig. 4 Visualization of codewords of brain connectivity dictionary obtained by Fisher Vector encoding of MADs with $p = 40$ in HCP dataset. Codewords are visualized using BrainNet Viewer (Xia et al. 2013), and depicted in an order such that Codeword1 and Codeword20 represent the codeword of GHE and GLE, respectively

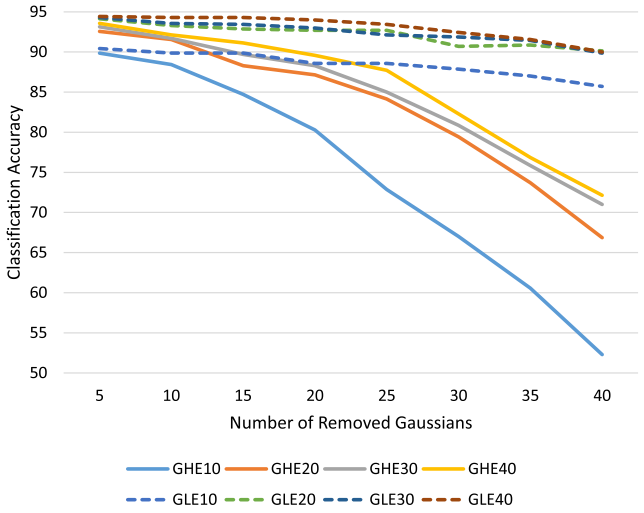

Fig. 5 Analysis of effect of removing Gaussians from representations on classification accuracy. Results of removal of Gaussians, whose FV entries have high energy (GHE), are plotted with solid lines, while results of removal of Gaussians, whose FV entries have low energy (GLE), are plotted with dashed lines. Removal of GHE decreases the accuracy more compared to removal of GLE

the occipital lobe. On the other hand, codeword of GLE has large values in central structures including Caudate, Putamen and Thalamus.

Simonyan et al. (2013) showed that GHEs represent the facial features whereas GLEs cover the background areas. In other words, they stated that GHEs are significant to compare human face images. Following our results, we conjecture that GHEs convey task discriminative information. In order to validate this assumption, and investigate the relationship between the energy of FV columns of Gaussians and the classification accuracy, we employ an energy based feature selection scheme for FV encoding. For this purpose, we select and remove columns from FVs, and perform classification. When we remove FV columns with minimum and maximum energy, we observe that the accuracy decreases slightly and significantly, respectively (see Fig. 5). We conclude that GHEs can improve task discrimination capacity of classifiers more compared to GLEs.

Discussion and conclusion

In this study, we propose a novel framework to encode local connectivity patterns of fMRI data for classification of cognitive tasks and states using the HCP and the EMR datasets. fMRI connectivity patterns are modeled by a set of local meshes formed around each node. MADs are estimated using ridge regression, assuming a linear relationship among nodes. Several encoding methods, such as FV, VLAD and BoW, are used to encode MADs. We also generate a *brain connectivity dictionary* by fitting a GMM to the set of all MADs to analyze the connectivity patterns of fMRI data with respect to anatomical regions.

FV encoding improves the performance of models which employ MADs. These results show that, representing samples using a function of distance of their features to codewords of the *brain connectivity dictionary* improves discrimination of cognitive states and tasks, compared to using only MAD features which represent local connectivity patterns of the samples. Moreover, we observe that FV encoding of MADs is more successful compared to VLAD and BoW methods. These results suggest that employing second order information extracted from features using FV provides more information compared to the Euclidean distance employed in VLAD and BoW encoding.

We also observe that statistical decorrelation of MADs obtained using PCA is a significant step for FV encoding. We conjecture that fitting GMMs to decorrelated MADs would also result in codewords representing diverse connectivity patterns. If MADs are not decorrelated using PCA, then classification performances decrease significantly for all encoding methods.

We further analyze task discrimination power of Gaussians obtained from FV encoding. We observe the relationship between the energy of FV columns corresponding to Gaussians and classification accuracy. Following that result, we visualize codewords of *brain connectivity dictionary* on a human brain template according to their discriminative power measured by an energy function.

Brain connectivity dictionary proposed in this study first provides the common connectivity patterns observed among brain nodes during cognitive states and during cognitive tasks. Contrary to brain activity maps that represent different levels of activations obtained in brain regions, *brain connectivity dictionary* provides us information about the frequently observed connectivity patterns obtained within a neighborhood. Second, *brain connectivity dictionary* allows us to generate a new representation of our samples. When we represent our samples employing their distance to these dictionary elements, we obtain the highest classification performance compared to the other methods as shown in Tables 3 and 5. Third, the connectivity patterns which provide the most significant information used to discriminate cognitive states can be recognized by analyzing energy values of Fisher Vector representations of each dictionary element.

As a future work, we plan to employ the proposed framework to further analyze the relationship between different tasks and cognitive processes by visualizing feature maps of MADs and their encoding on the AAL brain atlas.

Acknowledgments This work was completed when Itir Onal Ertugrul was with the Department of Computer Engineering, METU.

Funding This work was supported by CREST, JST, Grant Number JPMJCR14D1, the ImPACT Program of the Council for Science, Technology, and Innovation (Cabinet Office, Government of Japan) and TUBITAK Project No 116E091. Itir Onal Ertugrul was supported by TUBITAK 2211E.

Compliance with Ethical Standards

Conflict of interests Itir Onal Ertugrul, Mete Ozay and Fatos T. Yarman Vural declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors. Data used in this study were previously collected.

References

Abbas, A., Deligiannis, N., Andreopoulos, Y. (2015). Vectors of locally aggregated centers for compact video representation. In *International conference on multimedia and expo (ICME)* (pp. 1–6). IEEE.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., et al. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Behroozi, M., & Daliri, M.R. (2014). Predicting brain states associated with object categories from fMRI data. *Journal of Integrative Neuroscience*, *13*(04), 645–667.

Behroozi, M., & Daliri, M.R. (2015). Rdlpfc area of the brain encodes sentence polarity: a study using fMRI. *Brain Imaging and Behavior*, *9*(2), 178–189.

Cai, S., Chong, T., Peng, Y., Shen, W., Li, J., von Deneen, K.M., Huang, L., Initiative, A.D.N., et al. (2017). Altered functional brain networks in amnestic mild cognitive impairment: a resting-state fMRI study. *Brain Imaging and Behavior*, *11*(3), 619–631.

Carvajal, J., McCool, C., Lovell, B., Sanderson, C. (2016). Joint recognition and segmentation of actions via probabilistic integration of spatio-temporal fisher vectors. arXiv:160201601.

Chen, M., Han, J., Hu, X., Jiang, X., Guo, L., Liu, T. (2014). Survey of encoding and decoding of visual stimulus via fMRI: an image analysis perspective. *Brain Imaging and Behavior*, *8*(1), 7–23.

Daliri, M.R. (2012). Predicting the cognitive states of the subjects in functional magnetic resonance imaging signals using the combination of feature selection strategies. *Brain Topography*, *25*(2), 129–135.

Daliri, M.R. (2014). A hybrid method for the decoding of spatial attention using the meg brain signals. *Biomedical Signal Processing and Control*, *10*, 308–312.

Delhumeau, J., Gosselin, P.H., Jégou, H., Pérez, P. (2013). Revisiting the vlad image representation. In *International conference on multimedia* (pp. 653–656). ACM.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(Jan), 1–30.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2429.

Jaakkola, T.S., Haussler, D., et al. (1999). Exploiting generative models in discriminative classifiers. *Advances in Neural Information Processing Systems*, 487–493.

Jégou, H., Douze, M., Schmid, C., Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *Computer Vision and pattern recognition (CVPR)* (pp. 3304–3311). IEEE.

Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and alzheimer's disease. *Brain Imaging and Behavior*, *10*(3), 799–817.

Lee, Y.S., Peelle, J.E., Kraemer, D., Lloyd, S., Granger, R. (2015). Multivariate sensitivity to voice during auditory categorization. *Journal of Neurophysiology*, *114*(3), 1819–1826.

Liu, L., Shen, C., Wang, L., Van Den Hengel, A., Wang, C. (2014). Encoding high dimensional local features by sparse coding based fisher vectors. In *Advances in neural information processing systems* (pp. 1143–1151).

Meriño, L.M., Meng, J., Gordon, S., Lance, B.J., Johnson, T., Paul, V., Robbins, K., Vettel, J.M., Huang, Y. (2013). A bag-of-words model for task-load prediction from eeg in complex environments. In *International Conference on acoustics, speech and signal processing (ICASSP)* (pp. 1227–1231). IEEE.

Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*(1–2), 145–175.

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195.

Mızrak, E., Singmann, H., Öztekin, I. (2017). Forgetting emotional material in working memory. *Social Cognitive and Affective Neuroscience*, *1*, 10.

Onal, I., Ozay, M., Vural, F.T.Y. (2015a). Functional mesh model with temporal measurements for brain decoding. In *International Conference of the engineering in medicine and biology society (EMBC)* (pp. 2624–2628). IEEE.

Onal, I., Ozay, M., Vural, F.T.Y. (2015b). Modeling voxel connectivity for brain decoding. In: *International Workshop on pattern recognition in NeuroImaging* (pp. 5–8). IEEE.

Onal, I., Ozay, M., Mizrak, E., Oztekin, I., Yarman-Vural, F. (2017). A new representation of fMRI signal by a set of local meshes for brain decoding. *IEEE Transactions on Signal and Information Processing over Networks*.

Oneata, D., Verbeek, J., Schmid, C. (2013). Action and event recognition with fisher vectors on a compact feature set. In *International Conference on computer vision* (pp. 1817–1824). IEEE.

Ozay, M., Öztekin, I., Öztekin, U., Yarman-Vural, F.T. (2012). Mesh learning for classifying cognitive processes. CoRR arXiv:1205.2382.

Perronnin, F., & Dance, C. (2007). Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and pattern recognition (CVPR)* (pp. 1–8). IEEE.

Perronnin, F., Sánchez, J., Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. *Computer Vision–ECCV*, *2010*, 143–156.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Ville, D.V.D. (2011). Decoding brain states from fMRI connectivity graphs. *NeuroImage*, *56*(2), 616–626.

Rissman, J., Chow, T.E., Reggente, N., Wagner, A.D. (2016). Decoding fMRI signatures of real-world autobiographical memory retrieval. *Journal of Cognitive Neuroscience*, *28*(4), 604–620.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. *Cerebral Cortex*, *26*(6), 2563–2573.

Sánchez, J., & Redolfi, J. (2015). Exponential family fisher vector for image classification. *Pattern Recognition Letters*, *59*, 26–32.

Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J. (2013). Image classification with the fisher vector: theory and practice. *International Journal of Computer Vision*, *105*(3), 222–245.

Savelonas, M.A., Pratikakis, I., Sfikas, K. (2016). Fisher encoding of differential fast point feature histograms for partial 3d object retrieval. *Pattern Recognition*, *55*, 114–124.

Sekma, M., Mejdoub, M., Amar, C.B. (2015). Human action recognition based on multi-layer fisher vector encoding method. *Pattern Recognition Letters*, *65*, 37–43.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D. (2011). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*.

Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A. (2013). Fisher vector faces in the wild. In *British Machine vision conference* (Vol. 5, pp. 11).

Solmaz, B., Dey, S., Rao, A.R., Shah, M. (2012). Adhd classification using bag of words approach on network features. In *SPIE Medical Imaging* (pp. 83,144T–83,144T).

Sucu, G., Akbas, E., Oztekin, I., Mizrak, E., Vural, F.Y. (2016). Decoding cognitive states using the bag of words model on fMRI time series. In *Signal Processing and communication application conference (SIU)* (pp. 2245–2248). IEEE.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, *15*(1), 273–289.

Wang, J., Liu, P., She, M.F., Nahavandi, S., Kouzani, A. (2013). Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control*, *8*(6), 634–644.

Xia, M., Wang, J., He, Y. (2013). Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, *8*(7), e68,910.

Zhou, L., Wang, L., Liu, L., Ogunbona, P., Shen, D. (2016). Learning discriminative Bayesian networks from high-dimensional continuous neuroimaging data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(11), 2269–2283.