Neuroinformatics (2019) 17:197–210
https://doi.org/10.1007/s12021-018-9394-9

ORIGINAL ARTICLE

Multi-Objective Cognitive Model: a Supervised Approach for Multi-subject fMRI Analysis

Muhammad Yousefnezhad[1] · **Daoqiang Zhang**[1]

Published online: 9 August 2018

Abstract

In order to decode human brain, Multivariate Pattern (MVP) classification generates cognitive models by using functional Magnetic Resonance Imaging (fMRI) datasets. As a standard pipeline in the MVP analysis, brain patterns in multi-subject fMRI dataset must be mapped to a shared space and then a classification model is generated by employing the mapped patterns. However, the MVP models may not provide stable performance on a new fMRI dataset because the standard pipeline uses disjoint steps for generating these models. Indeed, each step in the pipeline includes an objective function with independent optimization approach, where the best solution of each step may not be optimum for the next steps. For tackling the mentioned issue, this paper introduces Multi-Objective Cognitive Model (MOCM) that utilizes an integrated objective function for MVP analysis rather than just using those disjoint steps. For solving the integrated problem, we proposed a customized multi-objective optimization approach, where all possible solutions are firstly generated, and then our method ranks and selects the robust solutions as the final results. Empirical studies confirm that the proposed method can generate superior performance in comparison with other techniques.

Keywords Multi-objective cognitive model · fMRI analysis · Multivariate pattern · Multi-objective optimization

Introduction

One of the primary goals in neuroscience is to understand how the neural activities in the human brain can be mapped to different cognitive tasks. Analyzing task-based functional Magnetic Resonance Imaging (fMRI) data is an interdisciplinary technique. Almost all supervised applications of fMRI analysis explicitly or implicitly employ Multivariate Pattern (MVP) algorithms for extracting and decoding brain patterns (Yousefnezhad and Zhang 2017b). In practice, MVP analysis can be formulated as a classification problem and predict patterns of neural responses, which are generated by distinctive cognitive tasks (Yousefnezhad and Zhang 2017a). In order to elaborate the brain mapping technique, imagine a subject watched two different categories of visual stimuli, including photos of cats and human faces, and we collected the neural activities in the form of fMRI dataset. Then, we have employed a subset of this data for training an MVP classification model in order to predict the categories (human face or cat) in the rest of stimuli (that are unseen in the training procedure). As the final product of an MVP analysis, decision surfaces are defined to distinguish the neural activities that belong to different categories of stimuli (Haxby et al. 2014). Decision surfaces can be used to understand mental diseases (Yousefnezhad and Zhang 2017a, b).

In practice, fMRI analysis is a challenging problem. Cognitive models generated by MVP methods must be validated by employing multi-subject fMRI images (Chen et al. 2015; Chen et al. 2016; Lorbert and Ramadge 2012; Yousefnezhad and Zhang 2017a). There are mainly two steps in an MVP standard pipeline, which must be applied to *preprocessed* fMRI images to generate the cognitive models, i.e., functional alignment, classification analysis (Yousefnezhad and Zhang 2017b). Further, each of these main steps can include some subtasks, such as applying feature selection before classification analysis (Chen

✉ Muhammad Yousefnezhad
myousefnezhad@nuaa.edu.cn

✉ Daoqiang Zhang
dqzhang@nuaa.edu.cn

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

et al. 2015; Yousefnezhad and Zhang 2017b). Different human brains naturally generate distinctive patterns (Yousefnezhad and Zhang 2017a; Haxby et al. 2011). The general assumption in the brain decoding is that the generated patterns are noisy 'rotation' of a shared space (Chen et al. 2015; Haxby et al. 2014; Lorbert and Ramadge 2012; Yousefnezhad and Zhang 2017a). Functional alignment seeks this space for mapping the neural activities before generating the cognitive models (Haxby et al. 2014). As one of the common subtask in functional alignment, nonlinear kernel functions are employed to improve the performance of the alignment (Lorbert and Ramadge 2012). The next step in the standard pipeline is employing binary classification methods, such as regularized Support Vector Machine (SVM) algorithm, for generating the cognitive models, i.e., decision surfaces (Chen et al. 2015; Haxby et al. 2014; Lorbert and Ramadge 2012; Mohr et al. 2015; Yousefnezhad and Zhang 2017a, b). Feature selection is a usual subtask, which must be applied before the classification analysis, to reduce the sparsity problem and increase the performance of the final model (Chen et al. 2014).

However, recent studies demonstrated that most of the models that generated by the standard pipeline cannot provide stable performance on the new fMRI datasets (Bennett et al. 2009; Chen et al. 2016; Eklund et al. 2016; Pauli et al. 2016). As discussed before, there are different steps (including the steps and subtasks) for MVP analysis. The main problem is that these steps employ disjoint objective functions, which are separately run to generate a cognitive model (Yousefnezhad and Zhang 2017b) and there is no unique solution for each of these objective functions (Chen et al. 2015). Therefore, an optimal result in one of these steps may not be optimum for the next steps. The problem gets worse when each step uses an independent optimization strategy, which cannot update the result of the previous steps based on the errors of current step to improve the performance of the MVP analysis. Indeed, this is a prevalent issue in the optimization problems, which is called dominant (Deb et al. 2002; Li et al. 2014, 2016; Zitzler and Künzli 2004). In other words, if we have distinctive solutions, the ideal solution for each step not only must be an optimal solution for that step but also it must be optimum for all of the next steps.

The main contributions of this paper are threefold: it firstly reformulates different steps of MVP pipeline as an integrated multi-objective problem, which is called Multi-Objective Cognitive Model (MOCM). We also introduce the novel concept of Intra-subject functional alignment, which can separately track the alignment error for each subject. Further, a customized optimization approach is developed for solving the integrated problem by incorporating the idea of non-dominated sorting into the multi-indicator algorithm. Indeed, non-dominated sorting seeks Pareto optimal solutions, and then indicators rank the robust solutions as the final results.

In this paper, "Background" briefly reviews some related works. Section "The Proposed Method" introduces the proposed method. Section "Experiments" reports empirical studies. Finally, "Discussions and Conclusions" presents conclusion and pointed out some future studies.

Background

Since task-based fMRI datasets can provide better spatial resolution in comparison with other modalities, most of the previous studies employed fMRI datasets in order to study human brains (Haxby et al. 2014). A crucial step in fMRI analysis is creating a model that is generalized across subjects (Chen et al. 2014, 2015, 2016; Hanke et al. 2014; Haxby et al. 2011; Xu et al. 2012; Yousefnezhad and Zhang 2017a). In other words, utilizing multi-subject fMRI data is necessary to validate the generated results across subjects (Haxby et al. 2011; Yousefnezhad and Zhang 2017a). However, functional neural images require precise alignment for boosting the performance of the final model (Chen et al. 2014; Haxby et al. 2011; Yousefnezhad and Zhang 2017a). In practice, they are two primary alignment approaches, including anatomical alignment and functional alignment, that must work in unison (Haxby et al. 2011; Xu et al. 2012). Indeed, anatomical alignment is a classical technique for preprocessing fMRI images. However, the performance of the anatomical alignment is limited based on the location, shape, and size of the functional loci (Rademacher et al. 1993; Watson et al. 1993). By contrast, functional alignment does not suffer the mentioned issues in the anatomical alignment (Haxby et al. 2011).

Hyperalignment (HA), as the most prevalent approaches for applying functional alignment, is an 'anatomy free' technique that can be written as a Canonical Correlation Analysis (CCA) problem (Haxby et al. 2014; Haxby et al. 2011; Yousefnezhad and Zhang 2017a). As discussed before, HA seeks a shared neural representation across subjects. The performance of MVP analysis by using functional alignment is significantly increased (Haxby et al. 2014; Haxby et al. 2011). Xu et al. proposed the Regularized Hyperalignment (RHA), which uses an EM algorithm to find the optimum template parameters iteratively (Xu et al. 2012). Lorbert et al. developed Kernel Hyperalignment (KHA) as a nonlinear extension of HA method (Lorbert and Ramadge 2012). Chen et al. introduced a two-phase method for functional alignment, which is called Singular Value Decomposition Hyperalignment (SVDHA). The neural activities are mapped in SVDHA to a low-dimensional space by using SVD, and then the mapped features are aligned by using HA techniques (Chen et al.

2014). Yousefnezhad et al. introduced Local Discriminant Hyperalignment (LDHA) as the first supervised HA method for MVP analysis (Yousefnezhad and Zhang 2017a).

MVP techniques utilize classification algorithms for predicting a new subject's neural activities. Cox et al. used both linear and nonlinear SVM algorithms for the human brain decoding (Cox and Savoy 2003). Norman et al. illustrated SVM superior performance to Gaussian Naive Bayes classifiers (Norman et al. 2006). Carroll et al. developed a new cognitive model by employing the Elastic Net (Zou and Hastie 2005) to predict and interpret the distributed neural responses with sparse models (Carroll et al. 2009). Mohr et al. analyzed different classification techniques, i.e., the first norm regularized SVM (Bradley and Mangasarian 1998), the second norm regularized SVM (Cortes and Vapnik 1995), the Elastic Net (Zou and Hastie 2005), and the Graph Net (Grosenick et al. 2013), to predict distinctive neural activities in the human brain (Mohr et al. 2015). They figured out the first norm regularized SVM can rapidly improve the classification performance in fMRI analysis (Mohr et al. 2015; Yousefnezhad and Zhang 2017b). Osher et al. developed a network-based method by employing the human brain's anatomical features in order to classify distinctive neural responses (Osher et al. 2015). Yousefnezhad et al. proposed two new ensemble learning approaches by utilizing weighted AdaBoost (Yousefnezhad and Zhang 2016), and Bagging (Yousefnezhad and Zhang 2017b).

Recent studies demonstrated that most of the generated models by the standard pipeline cannot provide stable performances on new fMRI datasets (Bennett et al. 2009; Chen et al. 2016; Eklund et al. 2016; Pauli et al. 2016). Some studies illustrated that results of General Linear Models (GLM) that are generated by different software packages (AFNI (Cox 1996), FSL (Jenkinson et al. 2012), and SPM (Penny et al. 2011)) on a specific problem can be highly unstable (Pauli et al. 2016). Since these linear models are utilized in most of fMRI analysis (i.e., MVP classification), unstable models rapidly decrease robustness of the final results (Bennett et al. 2009; Cai et al. 2016; Pauli et al. 2016). Eklund et al. proved that the cognitive models generated by rest-mode fMRI datasets for spatial extent could significantly increase inflated false-positive rates (Eklund et al. 2016). Chen et al. developed Convolutional Autoencoder (CAE) method for improving the stability of functional alignment to analyze the whole brain neural activities (Chen et al. 2016). Indeed, CAE employed Shared Response Model (SRM) (Chen et al. 2015) for functional alignment as well as the standard searchlight analysis (Guntupalli et al. 2016) for improving the stability of the generated cognitive model (Chen et al. 2015).

There are a few studies that used multi-objective optimization (Carroll et al. 2009; Kao 2009; Kao et al. 2012). Indeed, these approaches formulate different steps belonging to fMRI analysis by using multiple objective functions rather than using single objectives for each section. Then, these methods seek optimal solutions for all of the objective functions simultaneously. Here, we may seek multiple optimal solutions for a specific problem, where they must be ranked in order to find the best final result. There are two approaches for ranking better solutions that must work in unison, i.e., non-dominated sorting (Deb et al. 2002) and multi-indicator algorithm (Li et al. 2016). While non-dominated sorting seeks all possible solutions, indicators rank the robust solutions in each Pareto frontier as the final results. In fMRI studies, Kao proposed a multi-objective approach for estimating a general linear model between the design matrix and the neural activities (Kao 2009). Conroy et al. develop a multi-objective optimization for selecting models in fMRI analysis, where they provided a principled method to take into account both classification accuracy and stability (Conroy et al. 2013). In another study, Kao et al. utilized a modified version of Non-dominated Sorting Genetic Algorithm (NSGA-II) for generating the linear model between the fMRI responses and task-related events (Kao et al. 2012). Ma et al. developed a multi-objective MVP technique by using Hierarchical Heterogeneous Particle Swarm Optimization (HHPSO), where the classification problem is formulated as a binary SVM, and then HHPSO seeks optimal solutions (Ma et al. 2016).

The Proposed Method

As preprocessed fMRI dataset, $\mathbf{F}^{(i)} \in \mathbb{R}^{T \times V_{org}}, i = 1{:}S$ is defined, where S denotes the number of subjects, V_{org} is the number of voxels in the original space, T denotes the number of time points in units of Time of Repetitions (TRs). This paper assumes that the neural activities of each subject are column-wise standardized, i.e., $\mathbf{F}^{(i)} \sim N(0, 1)$. We can also consider this condition as a preprocessing step if the original data is not standardized. A linear model then can be formulated for each subject as follows:

$$\mathbf{F}^{(i)} = \mathbf{D}^{(i)}\beta^{(i)} + \varepsilon^{(i)}, i = 1{:}S \tag{1}$$

where $\mathbf{D}^{(i)} \in \mathbb{R}^{T \times C}$ denotes the design matrix, $\varepsilon^{(i)}$ is the error of estimation, $\beta^{(i)} \in \mathbb{R}^{C \times V_{org}}$ is the sets of regressors. In addition, C denotes the number of stimulus categories in the experiment (Yousefnezhad and Zhang 2016, 2017b). Design matrix can be generated by convolution of time samples (or onsets: $\tau^{(i)} \in \mathbb{R}^{T \times C}$) and the Hemodynamic Response Function (HRF) signal ($\mathcal{H}$), i.e., $\mathbf{D}^{(i)} = \tau^{(i)} * \mathcal{H}$ (Yousefnezhad and Zhang 2017b). Here, time synchronized stimulus ensures temporal alignment, including the $m\text{-}th$ time point for all of the subjects represents the same

simulation (Lorbert and Ramadge 2012; Xu et al. 2012; Yousefnezhad and Zhang 2017a). In order to estimate $\beta^{(i)}$, the first objective function is defined for tracking error ($\varepsilon^{(i)}$) in (1) as follows:

$$\theta_1 = \frac{1}{S}\sum_{i=1}^{S} \mathbf{F}^{(i)} - \mathbf{D}^{(i)}\beta^{(i)} \tag{2}$$

By considering (2), the stimuli in the training-set are considered time synchronized, i.e., each time point for all subjects illustrates the same simulation (Lorbert and Ramadge 2012; Xu et al. 2012). Consequently, the class labels for the training-set are defined by $\mathbf{Y} = \{y_m\}$, $y_m \in \{-1, +1\}, m = 1{:}T$. In order to generalize the proposed method, a mapping function can be defined as follows:

$$\mathbf{X}^{(\ell)} \in \mathbb{R}^{T\times V} = \Phi\left(\mathbf{D}^{(\ell)}\beta^{(\ell)}\right) \tag{3}$$

where $\Phi : \mathbb{R}^{T\times V_{org}} \rightarrow \mathbb{R}^{T\times V}$ can be considered for two different applications. It can be any kernel function (Lorbert and Ramadge 2012) that maps the voxels from original nonlinear space to a linear embedded space. Further, this function can be any feature selection/ranking function (Chen et al. 2014, 2015; Güçlü and Gerven 2015). In order to employ the original data, this function can be considered as a linear mapping, where $\Phi(\mathbf{x}) = \mathbf{x}$. We will analyze different applications of this function in the experiments section.

The next step is functional alignment. As mentioned before, the general assumption in the brain decoding is that the generated patterns in each brain are noisy 'rotation' of a shared space (Haxby et al. 2014; Yousefnezhad and Zhang 2017a; Chen et al. 2015; Lorbert and Ramadge 2012). Figure 1 illustrates an example for functional alignment in a two-voxel representation space. As depicted in this figure, Hyperalignment (HA) seeks a shared space by using the training-set, where the correlations between different stimuli are minimized. By considering (2), functional alignment can be defined as follows (Yousefnezhad and Zhang 2017a):

$$\theta_2 = \frac{1}{S}\sum_{\substack{i=1\\ j=i+1}}^{S} \left\|\mathbf{X}^{(i)}\mathbf{R}^{(i)} - \mathbf{X}^{(j)}\mathbf{R}^{(j)}\right\|_F^2$$
$$\text{s.t.} \quad \left(\mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)}\right)^{\top}\mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)} = \mathbf{I}, \qquad \ell = 1{:}S \tag{4}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{R}^{(\ell)} \in \mathbb{R}^{V\times V}$ denotes the mapping that must be calculated for each subject. Here, voxel correlation map $(\mathbf{X}^{(i)})^{\top}\mathbf{X}^{(j)}$, $i, j = 1{:}S$ in the most of fMRI studies is not full rank because the number of voxels is significantly more than TRs (Chen et al. 2014; Lorbert and Ramadge 2012; Xu et al. 2012; Yousefnezhad and Zhang 2017a). Since (4) must be calculated for any new subject in the testing-phase, it is not computationally efficient.

Lemma 3.1 *The* Eq. 4 *is equivalent to:*

$$\theta_2 = \frac{1}{S}\sum_{i=1}^{S} \left\|\mathbf{G} - \mathbf{X}^{(i)}\mathbf{R}^{(i)}\right\|_F^2$$
$$\text{s.t.} \quad \left(\mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)}\right)^{\top}\mathbf{X}^{(\ell)}\mathbf{R}^{(\ell)} = \mathbf{I}, \qquad \ell = 1{:}S \tag{5}$$

where $\mathbf{G} \in \mathbb{R}^{T\times V}$ *is the HA shared space:*

$$\mathbf{G} = \frac{1}{S}\sum_{j=1}^{S} \mathbf{X}^{(j)}\mathbf{R}^{(j)} \tag{6}$$

Proof In a nutshell, both (4) and (5) can be reformulated as $S^2 tr\left(\mathbf{G}^{\intercal}\mathbf{G}\right) - \left(S\sum_{i=1}^{S} tr\left(\left(\mathbf{X}^{(i)}\mathbf{R}^{(i)}\right)^{\intercal}\mathbf{X}^{(i)}\mathbf{R}^{(i)}\right)\right)$, where *tr()* denotes the trace function. Please see (Yousefnezhad and Zhang 2017a; Lorbert and Ramadge 2012) for details. □

$\mathbf{G}$ is called the HA shared space, which can be used for functional alignment in the testing-phase (Yousefnezhad and Zhang 2017a; Xu et al. 2012; Haxby et al. 2011; Chen et al. 2014).

Remark 3.1 The angle of rotation for all stimuli in each subject after mapping must be equal. This paper defines Intra-Subject Evaluation (ISE) as follows for calculating the angle of rotation for each category of stimuli:

$$ISE(\mathbf{x}, \mathbf{g}) = \frac{\mathbf{x}^{\top}\mathbf{g}}{\|\mathbf{x}\|\|\mathbf{g}\|} \tag{7}$$

where the vectors $\mathbf{x} \in \mathbb{R}^V$ and $\mathbf{g} \in \mathbb{R}^V$ respectively denote the neural activities in a specific time point before and after mapping.

By considering (7), the error of rotation for all subject is calculated as follows:

$$\theta_3 = \frac{1}{S}\sum_{\ell=1}^{S}\sum_{\substack{m=1\\ n=m+1}}^{C} \left\|\text{ISE}\left(\mathbf{x}_{m.}^{(\ell)}, \mathbf{g}_{m.}^{(\ell)}\right) - \text{ISE}\left(\mathbf{x}_{n.}^{(\ell)}, \mathbf{g}_{n.}^{(\ell)}\right)\right\|_F^2 \tag{8}$$

where row vector $\mathbf{x}_{m.}^{(\ell)} \in \mathbb{R}^V$ denotes the neural activities (all voxels) belong to *m-th* time point, i.e., $\mathbf{x}_{m.}^{(\ell)} = \{x_{mn}^{(\ell)} | x_{mn}^{(\ell)} \in$

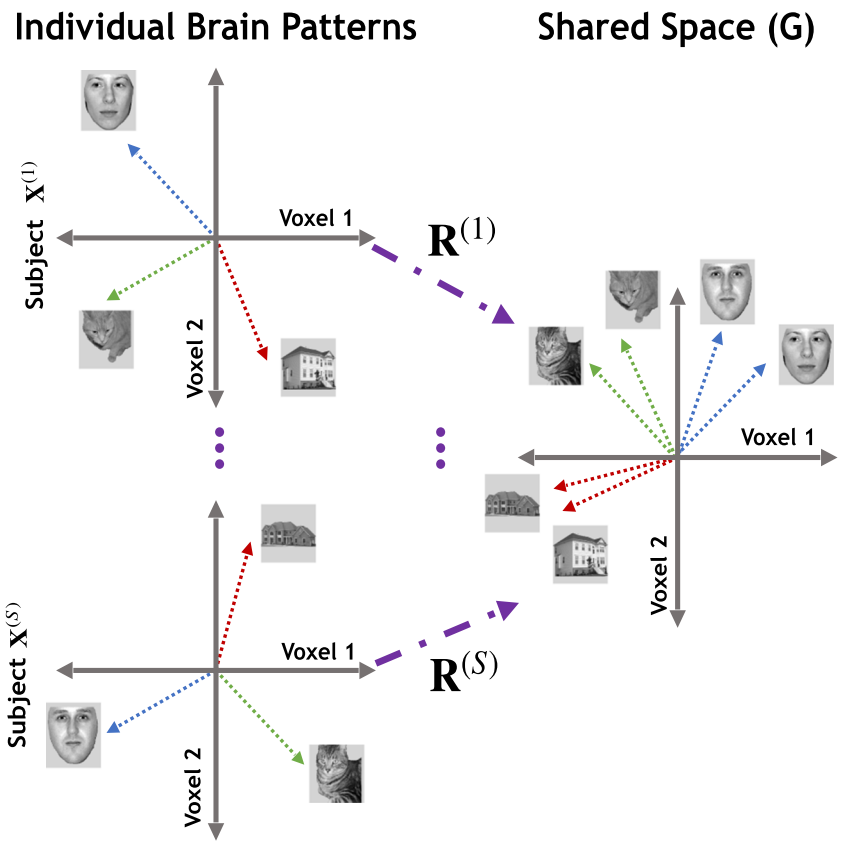


Fig. 1 An example of functional alignment in a two-voxel representation space. Here, the neural activities are depicted by vectors with different colors, and each color (i.e., blue, green, and red) represents a specific category of visual stimuli. Further, $R^{(\ell)}$ denotes the mapping from the original features to the shared space (Haxby et al. 2011; Lorbert and Ramadge 2012; Xu et al. 2012; Yousefnezhad and Zhang 2017a)

$\mathbf{X}^{(\ell)}, n = 1:V\}$. Further, we have the same notation for the shared space, $\mathbf{g}_{m.}^{(\ell)} \in \mathbb{R}^V = \{g_{mn}^{(\ell)} | g_{mn}^{(\ell)} \in \mathbf{G}^{(\ell)}, n = 1:V\}$.

Remark 3.2 As mentioned before, classification algorithms are employed in MVP analysis for generating the cognitive model. While we can use any algorithm for training a cognitive model, this paper employs L1 regularized SVM (Bradley and Mangasarian 1998) that is utilized in Mohr et al. (2015) as the best algorithm for fMRI analysis.

As the next step, a classification model is defined as follows:

$$\theta_4 = \frac{\alpha}{S} \sum_{i=1}^{S} \max\left(0, \mathbf{1}_T - \left(\text{diag}(\mathbf{Y})\mathbf{X}^{(i)}\mathbf{R}^{(i)}\mathbf{W}\right)\right)^2 + \|\mathbf{W}\|_1 \tag{9}$$

where the constraint $\alpha > 0$ is the SVM parameter, *diag* function create a square diagonal matrix from the class label vector $\mathbf{Y}$, $\mathbf{1}_T \in \mathbb{R}^T$ is ones vector, $\|.\|$ denotes the L1 norm, $\mathbf{W} \in \mathbb{R}^T$ is the decision surfaces for our cognitive model.

Training-phase for MOCM can be denoted by using following objective function:

$$\begin{aligned} &\min_{\mathbf{p}_{train}} \Theta_{train} \\ \text{s.t.} \quad &\Theta_{train} = K_{train}\left(\mathbf{F}, \tau; \mathbf{p}_{train}\right) \end{aligned} \tag{10}$$

where the vector $\Theta_{train} = [\theta_1, \theta_2, \theta_3, \theta_4]$ is the training error, the fMRI time series $\mathbf{F} = \{\mathbf{F}^{(i)}, i = 1{:}S\}$ and its corresponding onsets $\tau = \{\tau^{(i)}, i = 1{:}S\}$ are considered as the training-set, and the training parameters are defined by the vector $\mathbf{p}_{train} = \left[\beta^{(\ell)}, \mathbf{R}^{(\ell)}, \mathbf{W}\right], \ell = 1{:}S$. Here, K_{train} respectively employs (2), (4), (8), (9) in order to estimate $\theta_1, \theta_2, \theta_3$, and θ_4. In testing-phase, the following objective function is used:

$$\begin{aligned} &\min_{\mathbf{p}_{test}} \Theta_{test} \\ \text{s.t.} \quad &\Theta_{test} = K_{test}\left(\widehat{\mathbf{F}}, \widehat{\tau}, \mathbf{G}; \mathbf{p}_{test}\right) \end{aligned} \tag{11}$$

where the vector $\Theta_{test} = [\theta_1, \theta_2, \theta_3]$ is the error of testing-phase, the fMRI time series $\widehat{\mathbf{F}} = \{\widehat{\mathbf{F}}^{(i)}, i = 1{:}\widehat{S}\}$ and its corresponding onsets $\widehat{\tau}^{(i)} = \{\widehat{\tau}^{(i)}, i = 1{:}\widehat{S}\}$ denote the testing-set, $\widehat{S}$ is the number of subjects in the testing-set, $\mathbf{G}$ denotes the shared space that is calculated in the training-phase, and the testing parameters are defined by the vector $\mathbf{p}_{test} = \left[\widehat{\beta}^{(\ell)}, \widehat{\mathbf{R}}^{(\ell)}\right]$, $\ell = 1{:}S$. Here, K_{test} respectively uses (2), (5), (8) for estimating θ_1, θ_2, and θ_3. Furthermore, the final prediction can be generated by $\widehat{\mathbf{Y}}^{(i)} \in \mathbb{R}^T = \mathbf{1}_T - \left(\widehat{\mathbf{X}}^{(i)}\widehat{\mathbf{R}}^{(i)}\mathbf{W}\right)$ for all new subjects ($i = 1{:}\widehat{S}$), where $\mathbf{W}$ denotes the decision surfaces that is calculated in the training-phase.

Algorithm 1 Multi-Objective Cognitive Model (MOCM)

Input: $K(\mathbf{p})$ for training-phase or testing-phase, O as population size, $MaxIt$ as maximum iterations, $MaxSame$ as maximum iterations with the same optimal result.
Output: An optimal solution p_{opt}.

Method:
01. Random initial $\mathbf{P}^{(0)}$ as a set of the 1st O solutions.
02. $i = 0$, $j = 0$.
03. **While** $(i < MaxIt)$ and $(j < MaxSame)$ **Do**
04. $\mathbf{Q}^{(i)} = \emptyset$.
05. **Repeat** O-times **Do**
06. Randomly choose $\mathbf{p}_1, \mathbf{p}_2 \in \mathbf{P}^{(i)}$.
07. Create offspring $\mathbf{q} = (\mathbf{p}_1 + \mathbf{p}_2)/2$.
08. Update $\mathbf{Q}^{(i)} = \mathbf{Q}^{(i)} \cup \{\mathbf{q}\}$.
09. **End Repeat**
10. Random initial $\mathbf{E}^{(i)}$ with size O.
11. Generate new population: $\mathbf{U}^{(i)} = \mathbf{P}^{(i)} \cap \mathbf{Q}^{(i)} \cap \mathbf{E}^{(i)}$.
12. $\mathbf{P}^{(i+1)} = \mathbf{SORT}\left(K(\mathbf{p}), \mathbf{U}^{(i)}, O\right)$.
13. $\mathbf{p}_{opt}^{(i)}$ as the first sorted solution in $\mathbf{P}^{(i+1)}$.
14. **If** $\left(K(\mathbf{p}_{opt}^{(i)}) == K(\mathbf{p}_{opt}^{(i-1)})\right)$
15. $j = j + 1$.
16. **Else**
17. $j = 0$.
18. **End If**
19. $i = i + 1$.
20. **End While**
21. Return $\mathbf{p}_{opt}^{(i)}$.

Optimization

In the previous section, we introduced an integrated multi-objective function in order to apply supervised fMRI analysis. This section presents a customized multi-objective optimization approach for finding optimal solutions for both the training-phase and testing-phase. For simplicity, a generalized objective function is considered as follows:

$$\Theta \leftarrow K(\mathbf{p}) \tag{12}$$

where the function K and parameters $\mathbf{p}$ can be calculated by (10) for training-phase, and (11) is used for the testing-phase. Algorithm 1 depicts a general template as the optimization approach for MOCM. In this algorithm, O is the population size, $MaxIt$ denotes the maximum number of iterations, and $MaxSame$ is the maximum number of iterations with the same optimal solution. This algorithm firstly considers a set of O random solutions ($\mathbf{P}^{(0)}$) for the first iteration. Further, we create three different sets of solutions (with size O) in each iteration in order to generate a new population for the next iteration. As the first set, $\mathbf{P}^{(i)}$ is the best O solutions that are generated in the previous step. As the second set, we create O new offsprings by averaging randomly selected parents from $\mathbf{P}^{(i)}$. Indeed, this set tries to seek better solutions by combining the previous best solutions. As the last set, we create O new random solutions ($\mathbf{E}^{(i)}$) in order to increase the diversity of possible solutions. In fact, $\mathbf{E}^{(i)}$ can rapidly reduce the chance of the local optimum issue. These sets are combined as a new population ($\mathbf{U}^{(i)}$) with size $3O$, and then the **SORT()** function select the first O optimal solutions for the next step. Further, the first sorted solution ($\mathbf{p}_{opt}^{(i)}$) from $\mathbf{P}^{(i+1)}$ is considered as the best solution for $i\text{-}th$ iteration. As the finishing condition, the algorithm repeats $MaxIt$-times the optimization procedure unless the best solutions for $MaxSame$-times will be the same.

Algorithm 2 The **SORT** function

Input: Objective function $K(\mathbf{p})$, A set of solutions $\mathbf{U}$, Population size O

Output: $\mathbf{P}_{opt}$ as the first O optimal solutions.

Method:
01. **For Each** $\mathbf{p} \in \mathbf{U}$
02. $\Delta_p = \emptyset$, $n_p = 0$.
03. **For Each** $(\mathbf{q} \in \mathbf{U})$ **and** $(\mathbf{q} \neq \mathbf{p})$
04. $\Theta_p \leftarrow K(\mathbf{p})$, $\Theta_q \leftarrow K(\mathbf{q})$
05. **If** $(\Theta_p \prec \Theta_q)$
06. $\Delta_p = \Delta_p \cup \{\mathbf{q}\}$.
07. **Elsif** $\left(\Theta_q \prec \Theta_p\right)$
08. $n_p = n_p + 1$.
09. **End If**
10. **End For**
11. **If** $n_p = 0$ **Then**
12. $\Omega^{(1)} = \Omega^{(1)} \cup \{\mathbf{p}\}$.
13. **End If**
14. **End For**
15. $j = 1$, $\mathbf{P}_{opt} = \emptyset$.
16. **While** $(\Omega^{(j)} \neq \emptyset)$ **and** $(|\mathbf{P}_{opt}| < O)$
17. $\Omega^{(j+1)} = \emptyset$.
18. **I1**-evaluation: $a_p = \mathbf{I1}\left(\mathbf{p}, \Omega^{(j)}\right)$ for all $\mathbf{p} \in \Omega^{(j)}$.
19. **I2**-evaluation: $b_p = \mathbf{I2}\left(\mathbf{p}, \Omega^{(j)}\right)$ for all $\mathbf{p} \in \Omega^{(j)}$.
20. Order $\Omega^{(j)}$ (solutions with lowest max(a_p, b_p) in top).
21. **For Each** $(\mathbf{p} \in \Omega^{(j)})$
22. $\mathbf{P}_{opt} = \mathbf{P}_{opt} \cup \{\mathbf{p}\}$.
23. **For Each** $(\mathbf{q} \in \Delta_p)$
24. $n_q = n_q - 1$.
25. **If** $(n_q = 0)$
26. $\Omega^{(j+1)} = \Omega^{(j+1)} \cup \{\mathbf{q}\}$.
27. **End If**
28. **End For**
29. **End For**
30. $j = j + 1$.
31. **End While**

The key point in Algorithm 1 is the **SORT()** function. Algorithm 2 illustrates this function. As mentioned before, the optimization approach for MOCM is developed by incorporating the idea of non-dominated sorting (Deb et al. 2002) into the multi-indicator algorithm (Li et al. 2016). As the first step, the solutions in **U** are ranked based on the concept of domination.

Remark 3.3 The best solution ($\mathbf{p}_{opt}$) must dominate ($\prec$) (Deb et al. 2002; Li et al. 2016) all possible solutions in **U**. In other words, the estimation of the optimal result ($\Theta_{opt} \leftarrow K(\mathbf{p}_{opt})$) must satisfy the following conditions in comparison with all possible estimations ($\Theta_q \leftarrow K(\mathbf{q})$ for all $q \in \mathbf{U}$):

$$\begin{aligned} &\forall \theta_{opt}^{(j)} \in \Theta_{opt}, \theta_q^{(\ell)} \in \Theta_q \implies \theta_{opt}^{(j)} \leq \theta_q^{(\ell)} \\ &\exists \theta_{opt}^{(j)} \in \Theta_{opt}, \theta_q^{(\ell)} \in \Theta_q \implies \theta_{opt}^{(j)} < \theta_q^{(\ell)} \end{aligned} \tag{13}$$

In order to apply non-dominated sorting, Algorithm 2 firstly generates two criteria for ranking all possible solutions (**U**), i.e., the scale n_p and the matrix Δ_p for each solution. The scale n_p counts the number of solutions that can dominate the solution **p**, and the matrix Δ_p denotes the set of solutions that are dominated by the solution **p**. Further, the set of the first Front ($\Omega^{(1)}$) can be defined by the solutions that are not dominated by any solution ($\forall p \in \Omega^{(1)} \implies n_p = 0$).

As the second step, Algorithm 2 must create O optimal sorted solutions. As mentioned before, this paper uses the multi-indicator algorithm for evaluating error rates in the possible results. Indeed, these indicators can evaluate the robustness of the generated results. This paper employs two effective indicators, i.e., **I1** (Li et al. 2016; Zitzler and Künzli 2004) and **I2** (Li et al. 2016). As the first indicator, **I1** is defined as follows (Li et al. 2016; Zitzler and Künzli 2004):

$$\mathbf{I1}(\mathbf{q}, \mathbf{P}) = \sum_{\substack{\mathbf{p} \in \mathbf{P} \\ \mathbf{p} \neq \mathbf{q}}} \exp\left(\frac{-1}{0.05} \mathbf{I}_{\epsilon+}(\mathbf{q}, \mathbf{p})\right) \tag{14}$$

where $\mathbf{I}_{\epsilon+}(\mathbf{p}, \mathbf{q})$ is denoted as follows Zitzler and Künzli (2004):

$$\mathbf{I}_{\epsilon+}(\mathbf{p}, \mathbf{q}) = \min_{\epsilon} \left(\theta_\ell - \epsilon \leq \widehat{\theta}_\ell\right), \quad \begin{matrix} \forall \theta_\ell \in \Theta_p \leftarrow K(\mathbf{p}) \\ \forall \widehat{\theta}_\ell \in \Theta_q \leftarrow K(\mathbf{q}) \end{matrix} \tag{15}$$

Further, indicator **I2** is defined as follows Li et al. (2016):

$$\mathbf{I2}(\mathbf{q}, \mathbf{P}) = \min_{\substack{\mathbf{p} \in \mathbf{P} \\ \mathbf{p} \text{ precedes } \mathbf{q}}} (\mathbf{I}_{\mathbf{SDE}}(\mathbf{q}, \mathbf{p})) \tag{16}$$

where **p** precedes **q** means that the position (the original index) of **p** in the population **P** is smaller than the position **q** (Li et al. 2016). In addition, $\mathbf{I}_{\mathbf{SDE}}$ is calculated as follows Li et al. (2014):

$$\mathbf{I}_{\mathbf{SDE}}(\mathbf{p}, \mathbf{q}) = \left(\sum_{\ell} \mathrm{sd}\left(\theta_\ell, \widehat{\theta}_\ell\right)^2\right)^{\frac{1}{2}}, \quad \begin{matrix} \forall \theta_\ell \in \Theta_p \leftarrow K(\mathbf{p}) \\ \forall \widehat{\theta}_\ell \in \Theta_q \leftarrow K(\mathbf{q}) \end{matrix} \tag{17}$$

$$\mathrm{sd}(\theta, \widehat{\theta}) = \begin{cases} \theta - \widehat{\theta} & \text{if } \theta < \widehat{\theta} \\ 0 & otherwise. \end{cases} \tag{18}$$

In order to select the O optimal solutions, the set of $j\text{-}th$ Front solutions ($\Omega^{(j)}$) will be evaluated by **I1** and **I2**. Then, the elements of $\Omega^{(j)}$ will be ordered based on the evaluations, where the elements with lowest maximum error rates ($\max(a_p, b_p)$) are considered as the better solutions. Further, n_q for the solutions that are dominated by each of optimal solutions will be reduced, and if $n_q = 0$ then those solutions will be added to the set of Front solutions for the next step ($\Omega^{(j+1)}$). This procedure will be continued in order to select the O optimal solutions from **U**.

Figure 2 shows an example of MOCM solution, where two objective functions generate three different solutions (i.e., A, B, and C). The solutions A and B can dominate the solution C, including both $\theta_1(C)$ and $\theta_2(C)$ are greater than other solutions. However, we cannot select A or B based on non-dominated sorting because of $\theta_1(A) < \theta_1(B)$ and $\theta_2(A) > \theta_2(B)$. Therefore, the indicators **I1** and **I2** are employed to evaluate the solutions A and B. Here, A is selected as the optimal solution, where the maximum of indicator values ($\max(\mathbf{I1}_A, \mathbf{I2}_A) = 0.8$) in solution A is lower than B ($\max(\mathbf{I1}_B, \mathbf{I2}_B) = 0.9$).

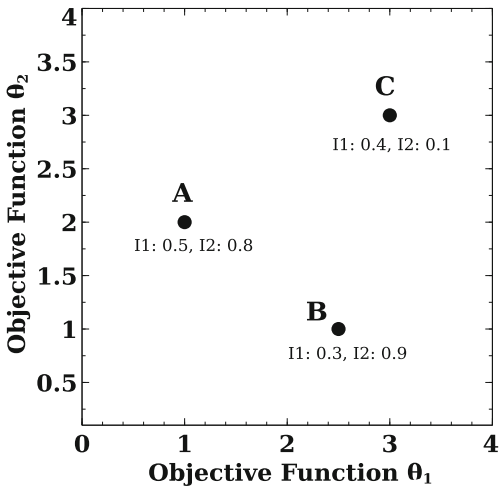


Fig. 2 An example of selecting optimal solution by using MOCM

Table 1 fMRI datasets

ID	Title	X	Y	Z	#	R	L	T	V	TR	TE	Scanner
DS005	Mixed-gambles task	53	63	52	16	48	2	240	450	2	30	Siemens 3 Tesla
DS105	Visual object recognition	79	95	79	6	71	8	121	1963	2.5	30	GE 3 Tesla
DS107	Word & object recognition	53	63	52	49	98	4	164	932	2	28	Siemens 3 Tesla
DS116	Auditory and visual oddball	53	63	40	17	102	2	170	2532	2	25	Philips 3 Tesla
DS117	Multi-subject, multi-modal	64	61	33	19	171	2	210	524	2	30	Siemens 3 Tesla
CMU	Meanings of nouns	51	61	23	9	9	12	360	17326	1	30	Siemens 3 Tesla

X, Y, Z as the size of 3D images; $\# = S + \widehat{S}$ denotes the number of subjects; R is the number of all runs (sessions); L denotes the number of stimulus categories; T is the number of time points; V denotes the number of voxels in ROI; TR is Time of Repetition in second; TE denotes Echo Time in millisecond

Experiments

Datasets

This paper utilizes 6 datasets, mostly shared by Open fMRI[1], for running empirical studies in this paper. These datasets are listed as follows:

- **DS005** includes 2 categories of risk tasks with the 50/50 chance of selection. In addition, Regions of Interest (ROI) is defined based on the original paper (Tom et al. 2007).
- **DS105** includes 8 categories of visual stimuli, i.e., gray-scale photos of cats, faces, houses, shoes, bottles, scissors, chairs, and scrambles (nonsense patterns). The neural activities in Ventral Temporal (VT) cortex is considered as the ROI in this dataset. Please refer to Haxby et al. (2011) and Haxby et al. (2014) for technical information.
- **DS107** contains 4 categories of visual stimuli, i.e., consonants, scrambles, objects, and words. The ROI is also defined based on the original study (Duncan et al. 2009).
- **DS117** includes MEG and fMRI images, where this paper just utilizes the fMRI data for running the empirical studies. Further, this dataset contains 2 categories of visual stimuli, i.e., human faces, and scrambles. In this dataset, the voxel responses in the VT cortex are considered as the ROI. Please see Wakeman and Henson (2015) for more information.
- **DS116** contains EEG signals and fMRI images. We just use the fMRI data in order to generate the experiments. This data includes 2 categories of audio and visual stimuli, including oddball tasks. Also, ROI is selected based on the original paper (Walz et al. 2013).
- **CMU** includes 12 semantic categories of word photos as the visual stimuli. Here, the ROI is defined based on the intersection of coordinates across subjects. Please refer to Mitchell et al. (2008) for more information.

Table 1 summarizes the technical information of these datasets. Further, this paper separately preprocessed all datasets by using FSL 5.0.9[2], i.e., slice timing, anatomical alignment, normalization, smoothing. Here, we have utilized the standard HRF signal generated by FSL in order to convolve the task events.

Performance Analysis

This section compares the performance of the proposed method with different MVP techniques. As a baseline, this paper reports the performance of L1 SVM (Bradley and Mangasarian 1998), which is used in Mohr et al. (2015) as the best algorithm for MVP analysis. Further, the performance of the original HA (Guntupalli et al. 2016; Haxby et al. 2011) and KHA (Lorbert and Ramadge 2012) are addressed for demonstrating the effect of functional alignment in MVP analysis. Here, KHA algorithm is applied by using the Gaussian kernel that introduced as the best kernel in the original paper (Lorbert and Ramadge 2012). Further, we utilized $1/n$ as the gamma parameter for all employed Gaussian kernels in this paper, where n is the number of features (Lorbert and Ramadge 2012; Yousefnezhad and Zhang 2017b). As a new graph-based method, the performance of Osher et al. method (Osher et al. 2015) is also reported in this paper. As a baseline for single-objective swarm optimization techniques in MVP analysis, the performance of PSO-SVM (Ma et al. 2016) is addressed in this paper. Moreover, the performance of HHPSO-SVM (Ma et al. 2016) and Kao et al. method (Kao et al. 2012) are reported as two multi-objective-based methods in MVP analysis. Finally, the performance of the MOCM method is addressed by using two different mapping functions, i.e., a linear mapping ($\Phi(\mathbf{x}) = \mathbf{x}$), and Gaussian kernel same as

[1] Available at http://openfmri.org

[2] Available at https://fsl.fmrib.ox.ac.uk

Table 2 Accuracy of classification methods (mean±std)

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117	CMU
L1 SVM (Mohr et al. 2015; Bradley and Mangasarian 1998)	71.65±4.97	85.29±3.49	81.25±3.62	69.24±3.28	76.61±2.73	73.62±3.15
L1 SVM + HA (Haxby et al. 2011; Guntupalli et al. 2016)	81.27±3.59	87.03±2.87	84.01±1.56	74.62±1.84	77.93±2.29	80.23±1.63
L1 SVM + KHA (Lorbert and Ramadge 2012)	83.06±2.36	90.05±2.39	86.68±1.71	80.51±2.12	84.22±1.44	83.49±2.03
Osher et al. (Osher et al. 2015)	84.55±2.02	90.82±1.87	85.62±1.95	78.91±2.04	86.81±1.79	85.01±1.97
PSO-SVM (Ma et al. 2016)	70.32±1.92	77.91±1.03	81.21±2.33	76.14±1.49	83.71±2.81	82.61±1.05
HHPSO-SVM (Ma et al. 2016)	90.17±1.01	94.46±1.23	89.91±1.67	96.03±0.56	**96.74±1.01**	87.62±1.03
Kao et al. (2012)	89.68±0.87	95.31±0.44	87.77±0.28	84.16±0.73	90.49±0.39	90.93±1.82
MOMVP (Linear kernel)	94.79±0.57	93.61±0.57	92.83±0.57	90.93±0.71	91.90±0.27	94.37±0.98
MOMVP (Gaussian kernel)	**96.10±0.29**	**98.34±0.29**	**96.79±0.59**	**97.09±0.33**	95.49±0.18	**96.02±0.92**

Best results are highlighted in each column

KHA method. In addition, there is no feature selection in this section. Like Ma et al. (2016), the population size is considered $O = 50$ for PSO, HHPSO, NSGA-II (in Kao et al. method), and MOCM. Moreover, we consider $MaxIt = 1000$ and $MaxSame = 5$ for all datasets. Like the previous studies (Lorbert and Ramadge 2012; Mohr et al. 2015; Yousefnezhad and Zhang 2016, 2017a, b), this paper firstly partitions the original fMRI images to training-set and testing-set by using leave-one-subject-out cross-validation. Then, the general linear model is calculated in the subject-level for all methods. After that, the functional alignment parameters are calculated for hyperalignment techniques. Next, this paper generates binary classifiers by applying one-versus-all strategy to the neural activities in the training sets. Finally, the performances of trained classifiers are evaluated by applying unseen testing sets and calculating the average of accuracy (or AUC) (Yousefnezhad and Zhang 2016, b). It is worth noting that the same structure and sample sets are applied to all evaluated methods in each iteration. Furthermore, the mentioned algorithms are implemented in the MATLAB R2016b (9.1) on a PC with certain specifications[3] by authors for generating the empirical studies.

Tables 2 and 3 respectively illustrate the classification Accuracy and Area Under the ROC Curve (AUC) in percentage (%). As depicted in these tables, L1 SVM cannot provide acceptable performance in comparison with other techniques because it just uses the anatomical alignment. Further, functional alignment techniques (HA and KHA) improved the performance of MVP analysis in comparison with L1 SVM method. In addition, HHPSO-SVM generated better results in comparison with PSO-SVM because it uses a multi-objective optimization approach. Moreover, the performance of Kao et al. method is significantly unstable because the optimization approach (NSGA-II) in this method cannot trace errors very well. Indeed, this is the main reason that we extend the indicators algorithm for improving the robustness of non-dominated sorting. Here, the performances of multi-objective approaches are more stable than the singular-objective methods (based on the standard deviation). Finally, the proposed method has generated better performance in comparison with other methods because it provided a robust and stable solution for MVP analysis by developing an integrated objective function and providing an effective optimization strategy. Indeed, the proposed method provides better performance when it is applied by using the Gaussian kernel that can map the nonlinear data points to a linear space. Furthermore, MOCM can calibrate the parameters generated in each step of fMRI analysis by tracing the errors in other steps. A good example is functional alignment techniques that can generate different solutions for a specific problem (Chen et al. 2015). While a single objective function generates these solutions, there is no way to rank or select one of them. However, MOCM can rank all possible solutions in each step of fMRI analysis (i.e., function aligning, classification, etc.) by tracing the effects of that solution on the other steps. Here, if we have two different alignment solutions for a specific problem, MOCM selects the solution with lowest classification error as the optimal solution. It is worth noting that we always select a set of optimal solutions in each iteration that has potential to generate better solutions in the next step (by creating new offsprings).

MVP Analysis by Using Feature Selection

This section analyzes the performance of MVP methods by using the features selection techniques. MOCM is compared with SVDHA (Chen et al. 2014), SRM (Chen et al. 2015), and CAE (Chen et al. 2016) as the state-of-the-art MVP methods that can apply feature selection before generating a cognitive model. Here, L1 SVM is used for generating

[3]DEL, CPU = Intel Xeon E5-2630 v3 (8×2.4 GHz), RAM = 64GB, OS = Ubuntu 16.04.2 LTS.

Table 3 Area under the ROC Curve (AUC) of classification methods (mean±std)

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117	CMU
L1 SVM (Mohr et al. 2015; Bradley and Mangasarian 1998)	68.37±4.01	80.91±3.21	80.72±2.88	66.85±3.05	72.12±1.48	70.08±2.94
L1 SVM + HA (Haxby et al. 2011; Guntupalli et al. 2016)	70.32±2.92	84.82±2.53	82.94±1.03	73.91±1.33	75.14±2.49	78.32±1.21
L1 SVM + KHA (Lorbert and Ramadge 2012)	82.22±2.42	88.81±1.61	83.36±2.12	77.41±1.97	81.54±1.92	81.76±2.92
Osher et al. (Osher et al. 2015)	81.83±2.86	89.54±1.74	82.02±2.43	75.08±1.12	84.08±1.84	82.27±2.06
PSO-SVM (Ma et al. 2016)	67.84±2.82	75.61±1.57	80.14±2.47	73.59±1.95	79.05±2.12	79.88±3.73
HHPSO-SVM (Ma et al. 2016)	87.91±1.83	92.39±1.73	86.12±0.99	**95.32±1.18**	90.73±1.59	87.01±1.61
Kao et al. (2012)	88.13±1.58	90.45±0.73	84.67±1.04	83.28±1.47	89.69±1.27	89.00±1.02
MOMVP (Linear kernel)	91.17±0.80	92.36±0.84	91.63±0.69	90.16±0.72	90.15±0.69	92.66±0.73
MOMVP (Gaussian kernel)	**94.37±0.63**	**97.71±0.58**	**93.22±0.49**	94.97±0.14	**93.72±0.31**	**95.79±0.42**

Best results are highlighted in each column

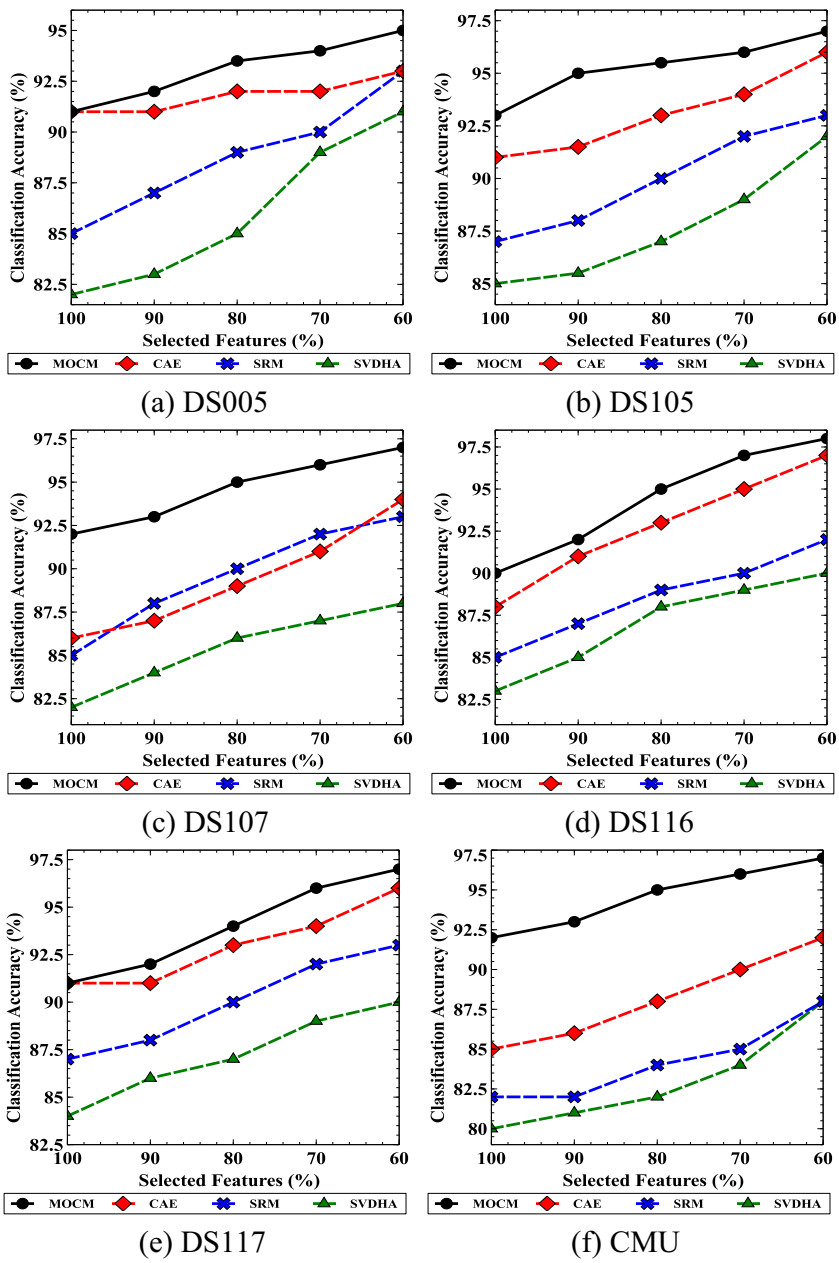
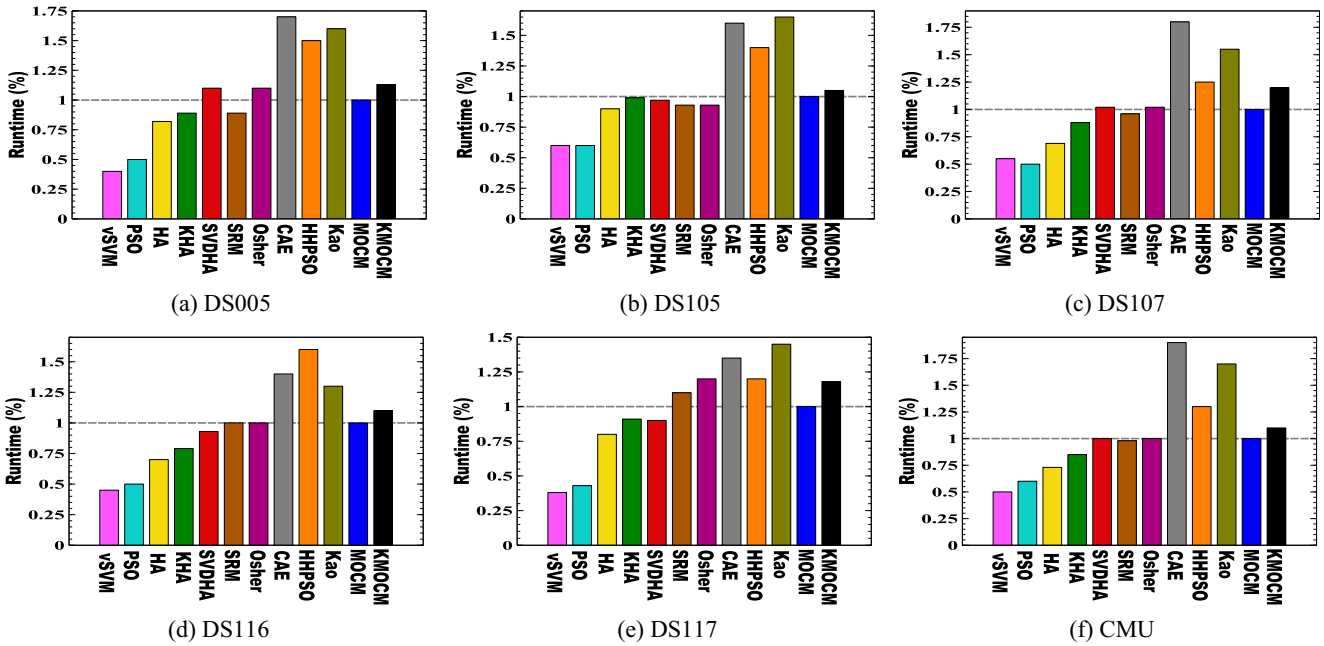


Fig. 3 MVP analysis by using feature selection

the cognitive models after each of the mentioned methods are applied on the preprocessed fMRI images for functional alignment. Like SVDHA, the proposed method employs a feature section function in terms of SVD analysis, where the mapping function $\Phi : V_{org} \rightarrow V, V_{org} \gg V$ is defined in order to generate the cognitive model (Chen et al. 2014). In other words, SVD decomposition is applied to the neural activities and then features are sorted based on the largest singular values. After that, we have selected the V features, where they have the largest V singular values in the decomposition. Next, the selected features are used in SVDHA and MOCM for training the classification model. For CAE method, the features are selected by reducing the number of units in the convolution neural network. In addition, we have selected features in SRM by changing the parameter k in this method, where k is the size of features for generating the mappings ($\mathbf{W}_i$) and the shared space ($\mathbf{S}$) in SRM method (Chen et al. 2015). It is worth noting that the feature selection procedure is applied separately to the training set and the testing set after these sets are partitioned by using cross-validation. Further, the setup of this experiment is same as the previous section (cross-validation, the population size, etc.). Figure 3 shows the performance of different methods by selecting 100% to 60% of features. As shown in this figure, the proposed method has generated better performance in comparison with other methods. Indeed, it can track errors of learning during the feature selection and then update different training coefficients ($\beta^{(i)}$, $\mathbf{R}^{(i)}$, $\mathbf{W}$) for minimizing the generated errors.

Runtime Analysis

This section analyzes runtime of different MVP methods. As mentioned before, all of the empirical studies are generated by using a specified PC. Figure 4 compares the runtime of MOCM with other functional alignment methods, where all runtime are scaled based on the proposed method (the runtime of MOCM is utilized as a unit). As this figure illustrates, there are four groups of methods based on the runtime. As the first group, SVM (Mohr et al. 2015; Bradley and Mangasarian 1998) and PSO-SVM (Ma et al. 2016) just employed a singular objective function and data without functional alignment. Therefore, they produce low accuracy (see previous sections) and runtime. As the second group, HA (Haxby et al. 2011), KHA (Lorbert and Ramadge 2012), SVDHA (Chen et al. 2014), Osher et al. method (Osher et al. 2015), SRM (Chen et al. 2015), and CAE (Chen et al. 2016) simultaneously utilized two singular objective functions for functional alignment and classification learning. Since HA, KHA, and SVDHA employed a single objective function for generating the function alignment parameters, i.e., the shared space ($\mathbf{G}$) and mapping functions ($\mathbf{R}^{(i)}$), the number of iteration for optimizing these parameters is naturally lower than a multi-objective solution. Thus, they are a little faster than MOCM. However, the performance of these methods are limited, and the optimization approaches in these methods cannot calibrate the alignment parameters based on the generated errors in GLM step or classification learning procedure. It is worth noting that the runtime of CAE is



(a) DS005 (b) DS105 (c) DS107

(d) DS116 (e) DS117 (f) CMU

Fig. 4 Runtime analysis

high because it employs deep learning method for aligning the neural activities. As the next group, HHPSO (Ma et al. 2016) and Kao et al. (2012) methods use the multi-objective approaches but just for the learning step. Indeed, these methods considered the functional alignment as the preprocessing step. By contrast, the proposed method does not need a separate step for functional alignment because it utilizes an integrated solution in order to apply the whole of procedures.

Discussions and Conclusions

As the final product of Multi-Objective Cognitive Model (MOCM), Fig. 5 depicts some examples of the generated cognitive models across categories of stimuli. Indeed, we visualized the decision surfaces ($\mathbf{W}$) that are generated in the training-phase. In order to create the cognitive model, we applied MOCM with a linear mapping ($\Phi(\mathbf{x}) = \mathbf{x}$) to the whole-brain fMRI images with the following parameters: $O = 50, MaxIt = 1000, MaxSame = 10$. This figure illustrates that different loci are activated based on distinctive stimuli. Further, the brain activities will be more focused in a certain region, when the stimuli just include the specific exemplars (such as human faces in DS117) rather than the abstract categories (concepts), e.g. objects in DS107. Indeed, this assumption is matched by the results of the previous studies (Haxby et al. 2011; Mohr et al. 2015; Mitchell et al. 2008; Yousefnezhad and Zhang 2016, 2017b), and can be considered as a shred of evidence for validating the generated model. It is worth noting that the proposed method can used for understanding how the human brain works and seeking new treatments for mental diseases.

There are several advantages to using multi-objective approach. Firstly, it can simplify the procedure of analysis. While other approaches need different steps with distinctive parameters (that may conflict with each other), we only need to apply a single step in the MOCM method for generating every thing, i.e., beta values, aligned features, and the classification model. The second advantage is tracing errors in different steps. Since we optimize a vector (i.e., the cost of different objective functions) at the same time rather than the disjoint single objective functions, we can rank and select possible solutions based on their generated errors in different steps. Since the set of optimal solutions (new offsprings) in each iteration are generated by using the ranked solutions of the previous iteration, they have potential to improve the quality of the final results in all steps simultaneously, including beta values, aligned features, classification models, etc.

In summary, this paper proposes MOCM as an integrated objective function in order to improve the performance and stability in the supervised fMRI analysis. By contrast of the previous methods, this objective function can apply both the functional alignment step and the learning step at the same time. Further, this objective function is generalized by using the kernel approach (for nonlinear data) and feature selection technique (for reducing the sparsity and noise). In order to solve the integrated objective function, a customized multi-objective optimization approach is

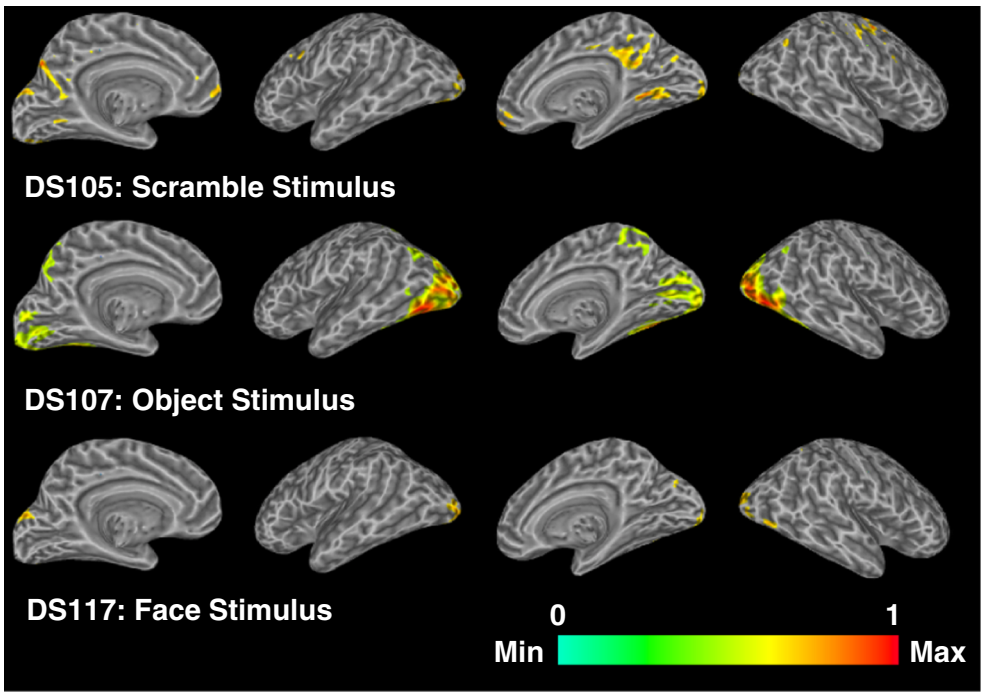


Fig. 5 Examples of cognitive models (decision surfaces), generated for each category of stimuli by applying MOCM to whole-brain datasets

developed by incorporating the idea of non-dominated sorting into the multi-indicator algorithm. Indeed, non-dominated sorting seeks all possible solutions, and then indicators rank the robust solutions as the final results. Empirical studies on multi-subject fMRI datasets confirm that the proposed method achieves superior performance to other state-of-the-art MVP techniques. In the future, we will plan to utilize the proposed method for improving the performance of other techniques in fMRI analysis, i.e unsupervised learning in RSA methods, multi-modality analysis, and neural hub detection.

Information Sharing Statement

The publicly available Open fMRI (RRID:SCR_005031) datasets are used in this paper that can be found via the GitHub site link: https://openfmri.org. Further, we have shared a preprocessed version of datasets in MATLAB format at https://easyfmridata.github.io. In addition, after this paper is accepted, the proposed method will be added to our GUI-based toolbox, i.e., available at https://easyfmri.github.io (RRID:SCR_016392).

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (61422204 and 61473149), and NUAA Fundamental Research Funds (NE2013105).

Compliance with Ethical Standards

Conflict of interests Muhammad Yousefnezhad and Daoqiang Zhang declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bennett, C.M., Baird, A., Miller, M.B., Wolfrod, G.L. (2009). Neural correlates of interspieces perspective taking in the post-mortem atlantic salmon: an argument for multiple comparisons correction. *Human Brain Mapping*, *1*, 1995.

Bradley, P.S., & Mangasarian, O.L. (1998). Feature selection via concave minimization and support vector machines. In *15th international conference on machine learning (ICML)* (Vol. 98, pp. 82–90). Association for computing machinery (ACM). Madison, Wisconsin, USA.

Cai, M.B., Schuck, N.W., Pillow, J.W., Niv, Y. (2016). A bayesian method for reducing bias in neural representational similarity analysis. In *Advances in neural information processing systems (NIPS)* (pp. 4951–4959).

Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., Rao, A.R. (2009). Prediction and interpretation of distributed neural activity with sparse models. *NeuroImage*, *44*(1), 112–122.

Chen, P.H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., Ramadge, P.J. (2015). A reduced-dimension fmri shared response model. In *28th advances in neural information processing systems (NIPS-15)* (pp. 460–468). Advances in neural information processing systems (NIPS). Montréal, Canada.

Chen, P.H., Guntupalli, J.S., Haxby, J.V., Ramadge, P.J. (2014). Joint svd-hyperalignment for multi-subject fmri data alignment. In *24th international workshop on machine learning for signal processing (MLSP)* (pp. 1–6). Reims, France: IEEE.

Chen, P.H., Zhu, X., Zhang, H., Turek, J.S., Chen, J., Willke, T.L., Hasson, U., Ramadge, P.J. (2016). A convolutional autoencoder for multi-subject fmri data aggregation. In *29th workshop of representation learning in artificial and biological neural networks. Advances in neural information processing systems (NIPS)*. Barcelona, Spain.

Conroy, B.R., Walz, J.M., Sajda, P. (2013). Fast bootstrapping and permutation testing for assessing reproducibility and interpretability of multivariate fmri decoding models. *PLoS ONE*, *8*(11), 1–11.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Cox, D.D., & Savoy, R.L. (2003). Functional magnetic resonance imaging (fmri) "brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Cox, R.W. (1996). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162–173.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

Duncan, K.J., Pattamadilok, C., Knierim, I., Devlin, J.T. (2009). Consistency and variability in functional localisers. *NeuroImage*, *46*(4), 1018–1026.

Eklund, A., Nichols, T.E., Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences (PNAS)*, *113*(28), 7900–7905.

Grosenick, L., Klingenberg, B., Katovich, K., Knutson, B., Taylor, J.E. (2013). Interpretable whole-brain prediction analysis with graphnet. *NeuroImage*, *72*, 304–321.

Güçlü, U., & van Gerven, M.A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10,005–10,014.

Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J.V. (2016). A model of representational spaces in human cortex. *Cerebral Cortex*, *26*(6), 2919–2934.

Hanke, M., Baumgartner, F.J., Ibe, P., Kaule, F.R., Pollmann, S., Speck, O., Zinke, W., Stadler, J. (2014). A high-resolution 7-tesla fmri dataset from complex natural stimulation with an audio movie. *Scientific Data*, *1*, 1–8.

Haxby, J.V., Connolly, A.C., Guntupalli, J.S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M. (2012). Fsl. *Neuroimage*, *62*(2), 782–790.

Kao, M.H. (2009). Multi-objective optimal experimental designs for er-fmri using matlab. *Journal of Statistical Software*, *30* (11), 1–13.

Kao, M.H., Mandal, A., Stufken, J. (2012). Constrained multiobjective designs for functional magnetic resonance imaging experiments via a modified non-dominated sorting genetic algorithm. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *61*(4), 515–534.

Li, B., Tang, K., Li, J., Yao, X. (2016). Stochastic ranking algorithm for many-objective optimization based on multiple indicators. *IEEE Transactions on Evolutionary Computation*, *20*(6), 924–938.

Li, M., Yang, S., Liu, X. (2014). Shift-based density estimation for pareto-based algorithms in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, *18*(3), 348–365.

Lorbert, A., & Ramadge, P.J. (2012). Kernel hyperalignment. In *25th advances in neural information processing systems (NIPS-12). Advances in neural Information Processing Systems (NIPS)* (pp. 1790–1798). Harveys, Lake Tahoe.

Ma, X., Chou, C.A., Sayama, H., Chaovalitwongse, W.A. (2016). Brain response pattern identification of fMRI data using a particle swarm optimization-based approach. *Brain Informatics*, *3*(3), 181–192.

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.

Mohr, H., Wolfensteller, U., Frimmel, S., Ruge, H. (2015). Sparse regularization techniques provide novel insights into outcome integration processes. *NeuroImage*, *104*, 163–176.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

Osher, D.E., Saxe, R.R., Koldewyn, K., Gabrieli, J.D., Kanwisher, N., Saygin, Z.M. (2015). Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cerebral Cortex*, *26*(4), 1668–1683.

Pauli, R., Bowring, A., Reynolds, R., Chen, G., Nichols, T.E., Maumet, C. (2016). Exploring fmri results space: 31 variants of an fmri analysis in afni, fsl, and spm. *Frontiers in Neuroinformatics*, *10*, 1–6.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E. (2011). Statistical parametric mapping: the analysis of functional brain images. Academic Press. ISBN: 978-0-12-372560-8.

Rademacher, J., Caviness, V.S., Steinmetz, H., Galaburda, A. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, *3*(4), 313–329.

Tom, S.M., Fox, C.R., Trepel, C., Poldrack, R.A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518.

Wakeman, D.G., & Henson, R.N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*, *2*, 1–10.

Walz, J.M., Goldman, R.I., Carapezza, M., Muraskin, J., Brown, T.R., Sajda, P. (2013). Simultaneous eeg-fmri reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. *Journal of Neuroscience*, *33*(49), 19,212–19,222.

Watson, J.D., Myers, R., Frackowiak, R.S.J., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., Shipp, S., Zeki, S. (1993). Area v5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex*, *3*(2), 79–94.

Xu, H., Lorbert, A., Ramadge, P.J., Guntupalli, J.S., Haxby, J.V. (2012). Regularized hyperalignment of multi-set fmri data. In *Statistical signal processing workshop (SSP)* (pp. 229–232). Ann Arbor, USA: IEEE.

Yousefnezhad, M., & Zhang, D. (2016). Decoding visual stimuli in human brain by using anatomical pattern analysis on fmri images. In *8th international conference on brain inspired cognitive systems (BICS'16)* (pp. 47–57). Beijing: Springer.

Yousefnezhad, M., & Zhang, D. (2017). Local discriminant hyperalignment for multi-subject fmri data alignment. In *34th AAAI conference on artificial intelligence (AAAI-17). Association for the advancement of artificial intelligence (AAAI)*. San Francisco, California, USA.

Yousefnezhad, M., & Zhang, D. (2017). Multi-region neural representation: a novel model for decoding visual stimuli in human brains. In *17th SIAM international conference on data mininig (SDM-17). Society for industrial and applied mathematics (SIAM)*. Houston, Texas, USA.

Zitzler, E., & Künzli, S. (2004). Indicator-based selection in multiobjective search. In *International conference on parallel problem solving from nature* (pp. 832–842). Birmingham: Springer.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.