Constrained Decoupled Dictionary Learning of Task fMRI for Brain Health Biomarkers

Sreekrishna Ramakrishnapillai[1] SREEKRISHNA.RAMAKRISHNAPILLAI@PBRC.EDU
[1] *Biomedical Imaging Center, Pennington Biomedical Research Center, Baton Rouge, LA, USA*

Harris R. Lieberman[2] HARRIS.R.LIEBERMAN.CIV@MAIL.MIL
[2] *Military Nutrition Division, US Army Research Institute of Environmental Medicine (USARIEM), Natick, MA, USA*

Jennifer C. Rood[1] JENNIFER.ROOD@PBRC.EDU
Stefan Pasiakos[2] STEFAN.M.PASIAKOS.CIV@MAIL.MIL
Lydia Bazzano[3] LBAZZANO@TULANE.EDU
[3] *Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, LA, USA*

Kori Murray[1] KORI.MURRAY@PBRC.EDU
Preetham Shankapal[4] PREETHSHANKPAL@GMAIL.COM
[4] *GE Healthcare – MR Engineering, Bengaluru, India*

Owen Carmichael[1] OWEN.CARMICHAEL@PBRC.EDU

Editors: Under Review for MIDL 2022

Abstract

Sparse dictionary learning methods have shown promise for representing task-based functional magnetic resonance imaging (tfMRI) data in terms of coherent spatiotemporal patterns of brain activity. However, methods that leverage prior knowledge, provide space-efficient representations, and correlate well with relevant clinical and scientific variables have been elusive. We propose a constrained decoupled dictionary learning method that uses prior knowledge related to task paradigms and the spatial locations of task-related brain regions to estimate decoupled spatial and temporal patterns that represent the fMRI data. The method identifies a high percentage of the spatial and temporal patterns programmed into simulated fMRI data. When applied to two real-world fMRI data sets, the method: 1. automatically identifies temporal and spatial patterns that were known by neuroscientists to be task-related a priori but were not provided as inputs to the method; 2. provides measurements that differ significantly between differing tasks; and 3. provides measurements that correlate well with relevant health information (e.g., fasting glucose level) in support vector regression models.

Keywords: Task-fMRI, SVM, Prediction

1. Introduction

Task-based functional magnetic resonance imaging (tfMRI) is used to identify brain regions whose blood oxygenation level dependent (BOLD) signal differs between differing conditions during performance of cognitive and motor tasks. Applying the general linear model (GLM) (Monti, 2011) separately at each voxel, followed by multiple comparison correction,

is the most commonly applied approach to the problem, but this approach does not efficiently identify broader spatiotemporal patterns (spanning multiple voxels) that capture more complex functional relationships between brain regions. Data-driven approaches such as independent component analysis (ICA) (Calhoun et al.) and sparse learning have been increasingly used to identify such patterns (Zhang et al., 2015; Zhao et al.; Liu et al.), but these methods largely do not leverage prior knowledge about the time course of task conditions or the probable spatial locations of brain regions involved in task execution. Our prior work (Ramakrishnapillai et al.) leveraged this prior knowledge to estimate a sparse dictionary of spatiotemporal elements whose characteristic BOLD signal covaries with task conditions and is biased toward including brain regions known a priori to be engaged by the task. We showed that using such prior information can provide more concise and robust summaries of tfMRI spatiotemporal patterns than competing methods when applied to both synthetic and real-world tfMRI data.

However, this prior work had two drawbacks. First, the method estimated joint spatiotemporal patterns, each of which reflected both a characteristic time course of the BOLD signal and a spatial distribution of locations where that time course is represented in the data. As such, the method was constrained to provide an equal number of spatial and temporal patterns, even when the data was more accurately represented by a larger number of spatial than temporal patterns (or vice versa). Second, while the prior method successfully represented tfMRI data, it was not clear whether the estimated spatiotemporal patterns were useful for any real-world neuroscience application. For example, the identified spatiotemporal patterns were not shown to differ between differing tasks or relate to relevant clinical indicators of brain health. These limitations reduced the ability of our spatiotemporal pattern method to have a substantial impact on tfMRI research.

In this paper, we alleviate these prior deficiencies via a novel framework that uses constrained optimization to identify arbitrary numbers of decoupled temporal and spatial sparse patterns. We show that the algorithm correctly identifies temporal and spatial structures programmed into simulated fMRI data. When applied to two real-world fMRI data sets, the method provides measurements that differ significantly between differing tasks and correlate well with relevant clinical variables in support vector classification and regression models.

2. Constrained Decoupled Sparse Representation of fMRI Data

There are 3 stages in the proposed framework. In stage one, spatially concatenated tfMRI data is factorized into a group-wise temporal dictionary matrix and a set of scan-level spatial patterns, i.e. spatial locations where those temporal patterns are represented in each scan. During this stage spatial and temporal constraints are used to bias a subset of the temporal and spatial patterns toward the known time course of task conditions and toward brain regions known to be involved in the task. In the second stage, the scan-level spatial patterns are aggregated and factorized to derive a group-level spatial dictionary matrix (which effectively removes redundancies across scans from the scan-level spatial pattern matrix) and sparse loading coefficients (Liu et al.). In stage 3, the sparse loading coefficients are used as descriptive biomarkers to predict relevant clinical variables using

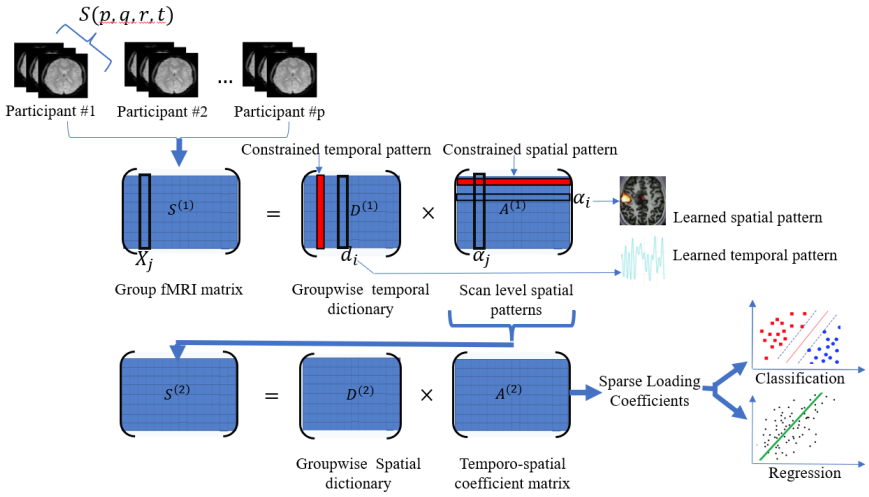

Figure 1: The proposed method organizes multiple tfMRI scans (top) into a matrix ($S^{(1)}$) and factorizes it into a matrix of canonical temporal patterns ($D^{(1)}$) and scan-level spatial patterns ($A^{(1)}$). The spatial patterns are aggregated across scans and factorized into a groupwise spatial dictionary ($D^{(2)}$) that removes redundancies among the scan-level spatial patterns, as well as sparse spatiotemporal coefficients ($A^{(2)}$) that quantify the degree to which individual scans represent each groupwise spatial dictionary element.

support vector classification or regression. These stages are described in the following sections.

2.1 Learning groupwise temporal dictionary and scan-level spatial patterns using constrained dictionary learning (Stage 1): In this stage, the set of tfMRI data sets is spatially concatenated into a matrix $S^{(1)}$ which is factorized into a group-wise temporal dictionary matrix $D^{(1)}$ containing time courses and a scan-level spatial pattern matrix $A^{(1)}$ using a constrained dictionary learning algorithm described in our previous work (Ramakrishnapillai et al.). Here, $S^{(1)} = D^{(1)} \times A^{(1)}$ where, $S^{(1)} = [S_1^{(1)}, S_2^{(1)} ..., S_p^{(1)}] \in \mathbb{R}^{t \times n \times p}$, $S_i^{(1)} \in \mathbb{R}^{t \times n}$, $D^{(1)} \in \mathbb{R}^{t \times K_1}$,$A^{(1)} = [A_1^{(1)}, A_2^{(1)} ..., A_p^{(1)}] \in \mathbb{R}^{K_1 \times n \times p}$ and $A_i \in \mathbb{R}^{K_1 \times n}$. Temporal constraints are incorporated into this factorization by changing the dictionary update step of the original algorithm (Mairal et al.) to a two-step process that updates the constrained and unconstrained temporal patterns separately. Specifically, let $\nabla(t)$ represent a time course of task conditions numerically, i.e., $\nabla(t) = 0$ (vs. 1) at time t when the first (vs. second) experimental condition is imposed. $\delta(t)$ denotes the temporal pattern obtained by convolving the task time course $\nabla(t)$ with the canonical hemodynamic response function. Given several δ, $\{\delta_1, \delta_2, \cdots \delta_m\}$ we encourage the discovery of temporal patterns that change in concert with changes in experimental conditions by constraining $D^{(1)}$ to the set of matrices W, each with members that contain columns similar to the δ:

$$W \triangleq \left\{ D^{(1)} \in \mathbb{R}^{t \times K} \;\middle|\; \begin{array}{l} \| \, d_j - \delta_j \, \|^2 \leq c_\delta, \;\; j = 1, \; \ldots, \; m \\ d_j^t d_j \leq 1, \;\; j = m+1, \; \ldots, \; K \end{array} \right\}$$

We also constrain a subset of the learned spatial patterns to be similar to those identified as task-engaged in previous fMRI studies, to encourage biological plausibility. For each vectorized spatial pattern α_i, which is a row in A, we generate P_i, a binary version of α_i that has a value of 1 wherever α_i is non-zero. We encourage any a priori spatial pattern M to have high spatial overlap with P_i by requiring that the Dice coefficient R between P_i and M be high:

$$R\left(P_i, M\right) = \frac{|\mathrm{P_i} \cap M|}{|M|} \tag{1}$$

The orthogonal matching pursuit algorithm in the sparse coding step of optimization is constrained to incorporate a new Dice-related constraint $\mathfrak{D}_\mathrm{c}$ on m designated rows of $A^{(1)}$. The columns of $A^{(1)}$ are also constrained, as in prior work (Ramakrishnapillai et al.) , to be sparse, i.e., to include no more than L nonzero pixels.

$$A^{(1)} \in \mathbb{R}^{K \times n} \; s.t \; \|\alpha_j\|_0^0 \leq L \; , \; R\left(P_i, M\right) \geq \mathfrak{D}_C, \quad i = 1, \; \ldots, \; m \tag{2}$$

$D^{(1)}$ is referred to as the temporal dictionary matrix, containing prototypical time courses representing the set of scans. The sparse coefficient matrix $A^{(1)}$ contains the scan-level spatial maps showing where each of those time courses are represented in each scan. The factorization was implemented on top of the base online optimization code within the SPAMS toolbox (Mairal et al., 2014).

2.2 Learning groupwise spatial dictionary and subject specific sparse loading coefficients (Stage 2): During the second stage, the scan-level spatial maps are factorized into a group-wise spatial dictionary matrix and sparse loading coefficients. Specifically, $S^{(2)} = D^{(2)} \times A^{(2)}$ where, $S^{(2)} = [(A_1^{(1)})^T, A_2^{(1)})^T ..., A_p^{(1)})^T] \in \mathbb{R}^{n \times (K_1 \times p)}$, $D^{(2)} \in \mathbb{R}^{n \times K_2}$ and $A^{(2)} = [C_1, C_2 ..., C_p] \in \mathbb{R}^{K_2 \times (K_1 \times p)}$ and $C_i \in \mathbb{R}^{K_2 \times K_1}$. $A^{(2}$ contains the sparse loading coefficient matrix for the scans, where C_i denotes the sparse loading coefficients corresponding to tfMRI scan i. This step removes redundant spatial patterns in $A^{(2}$ and thus $D^{(2}$ contains nonredundant spatial patterns common to all scans. The algorithm used to solve this factorization is the same as in Section 2.1 but without spatial and temporal constraints. The optimization was implemented in the SPAMS toolbox (Mairal et al., 2014) . $D^{(2}$ contains spatial maps of the distribution of time series representation across the set of scans.

2.3 Using sparse loading coefficients as biomarkers for classification and regression (Stage 3): The sparse loading coefficients output by Stage 2 can be used as biomarkers that summarize person-specific patterns of brain functioning. To assess the utility of these biomarkers, we train classifiers and regression machines to differentiate them between tasks and correlate them to relevant clinical variables. Specifically, for the classification problem, we have two tfMRI data sets per individual i: $S_i^{(1)} = [S_{i,t1}^{(1)}, S_{i,t2}^{(1)}]$ where $S_{i,y}$ is the tfMRI data set of the i^{th} subject completing the y^{th} task, $y \in \{t1, t2\}$. Task-specific sparse loading coefficients $C_{i,y}$ learned at the output of Stage 2 are used to train a Support Vector Machine (SVM) classification model.

For the regression problem participant i provides a tfMRI data set from a single task, which is run through the current method resulting in sparse loading coefficients C_i. Each participant also provides a set of brain-health-relevant clinical variables y_i. A support vector regression model is trained such that it fits a hyperplane $f(C) = C_i\beta + b$ for all C_i such that all residuals have a value less than a threshold ε i.e. $\forall i : |y_i - (C_i\beta + b)| \leq \varepsilon$ where β and b are the parameters of the hyperplane.

3. Experiments and Results

The algorithm was tested by applying it to simulated tfMRI data and determining how well its estimated spatial and temporal patterns represented patterns that were programmed into the simulated data *a priori*. The algorithm was also applied to two real tfMRI data sets to assess its ability to identify biologically-plausible spatial and temporal patterns, as well as generate sparse loading coefficients that differ significantly between tasks and correlate well with brain-health-related variables.

3.1 Testing Constrained Dictionary Learning (CDL) on Simulated fMRI Dataset

Generation of simulated fMRI dataset: SimTB (Erhardt et al., 2012) was used to generate synthetic task-related fMRI data sets. First, fifteen spatiotemporal patterns were provided to the simulator as ground truth, each consisting of a single spatial pattern defined on a 100 x 100 voxel image slice and a temporal pattern of 300 time points (TR = 2s). One of the temporal patterns (δ) corresponded to a ground-truth time course of changing task conditions convolved with a canonical hemodynamic response function. The corresponding spatial pattern was a single elliptical pattern covering a portion of the simulated brain region (Ramakrishnapillai et al.). The other 14 spatiotemporal patterns, representing various physiological sources of fMRI signal variability, were taken from a previous publication (Moreno et al.). Then, 20 synthetic datasets were generated by linearly combining the 14 sources using randomly generated amplitude scalings (with the scalings drawn uniformly from [0.5, 1]) along with the task related source. Translational head motion in the x and y directions was then drawn uniformly from the range of 0.02 to 0.5 times one voxel size and added to each synthetic scan. Finally, Gaussian noise at multiple amplitudes (SNR of 10, 20 and 50 dB) was added to create final synthetic datasets. We represented the ground-truth temporal patterns in vectorized form as the rows of a matrix, $D_g^{(1)}$, and represented the ground-truth spatial patterns in vectorized form as the columns of matrix $D_g^{(2)}$.

Performance evaluation: We evaluated the proposed algorithm in terms of similarity between rows and columns of estimated $D^{(1)}$ and $D^{(2)}$ matrices and those of $D_g^{(1)}$ and $D_g^{(2)}$. The temporal pattern was provided as a constraint to the optimizer. The spatial pattern M provided as a constraint to the optimizer is a binarized version of the elliptical spatial pattern corresponding to the constrained temporal source. For each run of each algorithm, $D^{(1)}, D^{(2)}$, $A^{(1)}$ and $A^{(2)}$ were initialized by randomly drawing entries from a normal distribution (zero mean and variance one) and normalizing columns to unit length. We completed differing runs with differing levels of sparsity L (see equation 2), which control the maximum number of nonzero elements in columns of A. Sparsity levels 2, 3, and 4 corresponded to 20%, 30% and 40% of the column being nonzero. Within each run, each paired step of optimization consisting of sparse coding and dictionary learning was executed 180 times. We completed 30 runs at each sparsity level for each tested algorithm, with each run applied to a different set of generated synthetic data. Each of the 15 learned temporal patterns in $D^{(1)}$ were correlated with each of the ground-truth time courses in $D_g^{(1)}$ via Pearson correlation. A ground-truth temporal pattern was considered recovered if there existed an estimated temporal pattern in $D^{(1)}$ whose Pearson correlation to it was greater than 0.9. A ground-truth spatial pattern was considered recovered if there existed a spatial pattern in $D^{(2)}$ that had 50 percent or more overlap with it. To calculate spatial

	Sparsity Level (L)	Recovered temporal Pattern	Recovered spatial pattern
SNR 10 dB	2	83.31	77.25
	3	82.56	75.23
	4	64.13	68.33
SNR 20 dB	2	91.82	88.51
	3	92.62	81.56
	4	92.30	86.58
SNR 50 dB	2	91.96	82.50
	3	92.27	88.63
	4	91.05	80.15

Table 1: Percentage of Recovered Temporal and Spatial patterns by using constrained decoupled dictionary learning on simulated fMRI data for various sparsity levels at different signal to noise ratios

overlap the spatial patterns were binarized first and the dice coefficient was calculated. The average percentage of recovered temporal and spatial patterns across all runs at a given sparsity level are shown in Table 1. The majority (between 75% and 92%) of temporal and spatial patterns were recovered correctly, the recovery rate did not vary substantially by the algorithm-prescribed sparsity level, and recovery rate degraded gracefully with reductions in SNR.

3.2 Real fMRI Datasets: All real fMRI scans were acquired on a 3T MRI scanner (General Electric, 750W Discovery, 32-channel quadrature head coil) using a blood oxygen level dependent echo-planar imaging (BOLD-EPI) pulse sequence at Pennington Biomedical Research Center. We used data from two different studies to assess the utility of the method for providing relevant biomarkers (see Sections 3.2.1 and 3.2.2 for data descriptions). All tfMRI data sets were preprocessed using Statistical Parametric Mapping 12 (SPM12) as described earlier (Carmichael et al., 2019). Briefly, preprocessing steps included slice-timing correction, head-motion correction, smoothing, coregistration to the T1-weighted image, and warping of the T1-weighted data to a standard coordinate frame. Each participant's T1 weighted anatomical image was segmented using SPM12, and the gray matter mask obtained as the result was warped to the MNI-ICBM152 template (Tzourio-Mazoyer et al., 2002). The gray matter mask was applied to the preprocessed functional images in the template space to isolate tfMRI signals originating from gray matter.

3.2.1 Classification of task type using SVM classifier

Experimental design: Fifty healthy young men aged 18 to 39 years completed fMRI scans during execution of the Attention Network Task (ANT) (Jennings et al., 2007) and the AX Continuous Performance Task (AX-CPT) (Lesh et al., 2013) as part of a clinical trial (Pasiakos et al., 2017). We used our method to estimate the sparse loading coefficients of $A^{(2)}$ from each scan with each task, and evaluated the coefficients as biomarkers of brain functional state. Specifically, we trained a support vector machine classifier to discriminate between sparse loading coefficients generated during ANT performance and those generated during AX-CPT performance. A sparsity level of 4 was used as it gave optimal results on

the synthetic data. In a leave-one out cross-validation (LOOCV) approach, we repeatedly trained the classifier using all but one set of sparse loading coefficients and assessed classifier performance on the remaining data point. The Python Scikit-Learn library (Pedregosa et al., 2011) was used for carrying out classification. For the SVM classification we employed linear, degree 3 polynomial, and radial basis function (RBF) kernels with the parameter $\gamma = 0.1$. The regularization parameter was set to the default value of 1. The data was not scaled as the coefficients were already scaled between 0 and 1. In addition, we compared estimated spatial patterns with ground truth spatial patterns derived from the literature (Jennings et al., 2007; Lesh et al., 2013), and confirmed that one of the estimated temporal patterns matched that of the time course of task conditions, which was provided to the algorithm as a loose constraint. The temporal patterns were compared by computing the Pearson correlation coefficient and spatial patterns were compared in terms of the dice coefficients. The ground truth time courses of task conditions corresponded to the executive control, alerting, and orienting contrasts when applied to the ANT data (Fan et al., 2005) (contrasts a, b and c respectively in Table 2) and the contrast between A cues and B cues when applied to AXCPT task (Lopez-Garcia et al., 2016)(contrast d in Table 2).

Results: Our constrained decoupled dictionary learning (CDDL) method identified spatial patterns known to be associated with the ANT and AX-CPT with higher fidelity than the previously presented, constrained dictionary learning (CDL) method (Table 2). The CDDL method also identified temporal patterns that were consistent with the ground truth task paradigms. The Pearson correlation coefficient (PCC) was above 0.80 for all contrast types for CDDL compared to values below 0.60 for CDL. The average PCC between the time courses of the identified task-evoked components and the task paradigms is 0.875. The support vector machine correctly classified sparse loading coefficients as arising from ANT vs. AXCPT with 92.67% accuracy on average across kernel types (Table 3).

	PCC		Overlap	
	DL	**CDL**	DL	**CDL**
a	0.53	**0.91**	0.59	**0.81**
b	0.58	**0.83**	0.61	**0.65**
c	0.47	**0.88**	0.51	**0.69**
d	0.56	**0.80**	0.61	**0.73**

Table 2: Temporal similarity in terms of Pearson correlation coefficient (PCC) and spatial overlap rate in terms of dice coefficient of all identified task evoked components for constrained Dictionary Learning (CDL) and Constrained Decoupled Dictionary Learning (CDDL)

Kernel	**Accuracy**
Linear	92
Polynomial	93
RBF	93

Table 3: Accuracy of SVM classifier in classifying the coefficients learned from two types of task fMRI data (ANT and AXCPT) for 3 different types of kernel functions

3.2.2 Prediction of clinical variables using SVM Regression:

Experimental design: Task fMRI data during completion of a Stroop task was collected from 100 participants in the Bogalusa Heart Study (Berenson, 2001). As above, our method was used to derive temporal patterns and sparse loading coefficients, which were then assessed for their plausibility and utility as biomarkers of brain health. A sparsity level of 4 was used here as well. To test the validity of groupwise temporal patterns we calculated the

	Glucose	HomA -IR	Insulin
Linear	40	41	45
Polynomial	38	45	42
RBF	39	44	43

Table 4: Mean Absolute Percentage Error (MAPE) of the Predicted cardiometabolic variables using SVM regression of various kernel types

Pearson correlation coefficient (PCC) between each of them and the task paradigm. To assess the utility of the sparse loading coefficients, we completed SVM based regression using the sparse loading coefficients as predictors and brain-health-relevant cardiometabolic variables, including fasting glucose level, homeostatic model assessment for insulin resistance (Homa-IR) index and fasting insulin level as outcomes. The Python Scikit-Learn library (Pedregosa et al., 2011) was used for carrying out regression using LOOCV scheme. We employed the same 3 kernel types described in subsection 3.2.1 for the SVM regression as well. The performance results for the algorithm are tabulated in terms of Mean Absolute Percentage Error (MAPE) in Table 4. From the table it can be inferred that for all kernel types, the SVM regressor was able to predict the cardiometabolic variables with MAPE less than 45. The prediction of glucose values achieved the least MAPE for all kernel types.

4. Conclusion

Combining prior knowledge, decoupling of spatial and temporal patterns, and high-dimensional classification and regression enabled this method to improve identification of clinically relevant biomarkers from real and synthetic fMRI data. Future work should further validate the sparse loading coefficients for biomarker applications, and assess the impact of algorithmic advances, such as advanced sparse regression methods and intelligent training data sampling, on algorithm performance.

Acknowledgments

The data used in the classification problem was taken from a project funded by Collaborative Research to Optimize Warfighter Nutrition projects I and II, and the Defense Health Program, Joint Program Committee-5, Military Operational Medicine Research Program. The data for regression analysis was taken from a project funded by NIH grants R01AG041200 and R01AG062309. Additional support provided by the Pennington Biomedical Research Foundation.

References

Gerald S Berenson. Bogalusa heart study: a long-term community study of a rural biracial (black/white) population. *The American journal of the medical sciences*, 322(5):267–274, 2001. ISSN 0002-9629.

Vince D Calhoun, Tulay Adali, Lars Kai Hansen, Jan Larsen, and James J Pekar. Ica of functional mri data: an overview. In *in Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*. Citeseer.

Owen Carmichael, Patrick Stuchlik, Sreekrishna Pillai, Geert-Jan Biessels, Ram Dhullipudi, Anna Madden-Rusnak, Shane Martin, Daniel S Hsia, Vivian Fonseca, and Lydia Bazzano. High-normal adolescent fasting plasma glucose is associated with poorer midlife brain health: Bogalusa heart study. *The Journal of Clinical Endocrinology Metabolism*, 104 (10):4492–4500, 2019. ISSN 0021-972X.

Erik B Erhardt, Elena A Allen, Yonghua Wei, Tom Eichele, and Vince D Calhoun. Simtb, a simulation toolbox for fmri data under a model of spatiotemporal separability. *Neuroimage*, 59(4):4160–4167, 2012. ISSN 1053-8119.

Jin Fan, Bruce D McCandliss, John Fossella, Jonathan I Flombaum, and Michael I Posner. The activation of attentional networks. *Neuroimage*, 26(2):471–479, 2005. ISSN 1053-8119.

Janine M Jennings, Dale Dagenbach, Christine M Engle, and Laura J Funke. Age-related changes and the attention network task: An examination of alerting, orienting, and executive function. *Aging, Neuropsychology, and Cognition*, 14(4):353–369, 2007. ISSN 1382-5585.

Tyler A Lesh, Andrew J Westphal, Tara A Niendam, Jong H Yoon, Michael J Minzenberg, J Daniel Ragland, Marjorie Solomon, and Cameron S Carter. Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. *NeuroImage: Clinical*, 2:590–599, 2013. ISSN 2213-1582.

Huan Liu, Mianzhi Zhang, Xintao Hu, Yudan Ren, Shu Zhang, Junwei Han, Lei Guo, and Tianming Liu. Fmri data classification based on hybrid temporal and spatial sparse representation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 957–960. IEEE. ISBN 1509011722.

Pilar Lopez-Garcia, Tyler A Lesh, Taylor Salo, Deanna M Barch, Angus W MacDonald, James M Gold, J Daniel Ragland, Milton Strauss, Steven M Silverstein, and Cameron S Carter. The neural circuitry supporting goal maintenance during cognitive control: a comparison of expectancy ax-cpt and dot probe expectancy paradigms. *Cognitive, Affective, Behavioral Neuroscience*, 16(1):164–175, 2016. ISSN 1531-135X.

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696. ACM. ISBN 1605585165.

Julien Mairal, F Bach, J Ponce, G Sapiro, R Jenatton, and G Obozinski. Spams: A sparse modeling software, v2. 3. *URL http://spams-devel. gforge. inria. fr/downloads. html*, 2014.

Martin M Monti. Statistical analysis of fmri time-series: a critical review of the glm approach. *Frontiers in human neuroscience*, 5:28, 2011. ISSN 1662-5161.

Manuel Morante Moreno, Yannis Kopsinis, Eleftherios Kofidis, Christos Chatzichristos, and Sergios Theodoridis. Assisted dictionary learning for fmri data analysis. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 806–810. IEEE. ISBN 1509041176.

Stefan M Pasiakos, Claire E Berryman, J Philip Karl, Harris R Lieberman, Jeb S Orr, Lee M Margolis, John A Caldwell, Andrew J Young, Monty A Montano, and William J Evans. Physiological and psychological effects of testosterone during severe energy deficit and recovery: a study protocol for a randomized, placebo-controlled trial for optimizing performance for soldiers (ops). *Contemporary clinical trials*, 58:47–57, 2017. ISSN 1551-7144.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011. ISSN 1532-4435.

Sreekrishna Ramakrishnapillai, Harris R Lieberman, Jennifer C Rood, Stefan M Pasiakos, Kori Murray, Preetham Shankapal, and Owen T Carmichael. Constrained learning of task-related and spatially-coherent dictionaries from task fmri data. In *International Workshop on Machine Learning in Clinical Neuroimaging*, pages 165–173. Springer.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002. ISSN 1053-8119.

Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. A survey of sparse representation: algorithms and applications. *IEEE access*, 3:490–530, 2015. ISSN 2169-3536.

Lin Zhao, Huan Liu, Xi Jiang, Shijie Zhao, Zhibin He, Tianming Liu, Lei Guo, and Tuo Zhang. A task performance-guided model of functional networks identification. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1590–1593. IEEE. ISBN 1538636417.