

Multi-study evaluation of neuroimaging-based prediction of medication class in mood disorders

Mustafa S. Salman^{a,b,*}, Eric Verner^a, H. Jeremy Bockholt^a, Zening Fu^a, Maria Misiura^a, Bradley T. Baker^a, Elizabeth Osuch^{c,d}, Jing Sui^{a,e}, Vince D. Calhoun^{a,b}

^a Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS) [Georgia State University, Georgia Institute of Technology, and Emory University], Atlanta, GA, USA

^b School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

^c Lawson Health Research Institute, London Health Sciences Centre, FEMAP, London, Ontario, Canada

^d Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

^e Institute of Automation, Chinese Academy of Sciences, and the University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:

Neuromark
Treatment response
Bipolar disorder
Major depressive disorder
Kernel SVM

ABSTRACT

Clinicians often face a dilemma in diagnosing bipolar disorder patients with complex symptoms who spend more time in a depressive state than a manic state. The current gold standard for such diagnosis, the Diagnostic and Statistical Manual (DSM), is not objectively grounded in pathophysiology. In such complex cases, relying solely on the DSM may result in misdiagnosis as major depressive disorder (MDD). A biologically-based classification algorithm that can accurately predict treatment response may help patients suffering from mood disorders. Here we used an algorithm to do so using neuroimaging data. We used the neuromark framework to learn a kernel function for support vector machine (SVM) on multiple feature subspaces. The neuromark framework achieves up to 95.45% accuracy, 0.90 sensitivity, and 0.92 specificity in predicting antidepressant (AD) vs. mood stabilizer (MS) response in patients. We incorporated two additional datasets to evaluate the generalizability of our approach. The trained algorithm achieved up to 89% accuracy, 0.88 sensitivity, and 0.89 specificity in predicting the DSM-based diagnosis on these datasets. We also translated the model to distinguish responders to treatment from nonresponders with up to 70% accuracy. This approach reveals multiple salient biomarkers of medication-class of response within mood disorders.

1. Introduction

Many studies have reported fundamental differences between major depressive disorder (MDD) and bipolar disorder (BD) (Osuch et al., 2018; de Almeida and Phillips, 2013; Bowden, 2005; Perlis et al., 2006). The complexity of symptoms exhibited by unipolar and bipolar disorder patients often leads to the wrong diagnosis and treatment. The current “gold standard” for such diagnosis, the Diagnostic and Statistical Manual (DSM), may lead to misdiagnosis as MDD without any evident symptoms of mania (American Psychiatric Association, 2013). BD patients tend to spend more time in depressive states, which may mislead clinicians to prescribe antidepressants (ADs) and worsen BD type I (Judd et al., 2002). On the other hand, mood stabilizers (MSs) may fail to treat MDD effectively. The patient’s recovery can be vastly improved with the correct mood diagnosis and selection of treatment. This study uses a

kernel support vector machine (SVM) classification algorithm in multi-dataset and multi-feature cases to predict medication-class of response (MS vs. AD) from fMRI data.

Functional magnetic resonance imaging (MRI) has been used in numerous studies to distinguish unipolar and bipolar disorders diagnosed by DSM. He et al. demonstrated that the striatum-precuneus connectivity estimated from fMRI could serve as a marker for differentiating these groups ($N = 84,50$ patients) (He et al., 2019). In their study, Rai et al. also showed the role of default mode and fronto-parietal network connectivity in distinguishing ($N = 116,77$ patients) (Rai et al., 2021). On the other hand, Han et al. found that the functional network switching rate is altered differently in BD and MDD ($N = 162, 101$ patients) (Han et al., 2020). These recent studies demonstrate the utility of fMRI data of small sample patients to understand these conditions. Furthermore, it validates our goal of investigating treatment

* Corresponding author at: School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

E-mail address: esalman@gatech.edu (M.S. Salman).

<https://doi.org/10.1016/j.pychresns.2023.111655>

Received 24 September 2022; Received in revised form 20 March 2023; Accepted 26 April 2023

Available online 9 May 2023

0925-4927/© 2023 Elsevier B.V. All rights reserved.

response to augment DSM-based diagnosis using the same modality (fMRI).

Group independent component analysis (ICA) is a popular and highly used among data-driven algorithms for multi-subject fMRI studies (Calhoun et al., 2001; McKeown et al., 1998). The spatial group ICA approach estimates spatial patterns of brain activity, or spatial maps (SMs), which are maximally spatially independent across subjects. In the subsequent back-reconstruction step, each subject's data is decomposed into unique time courses (TCs) and significantly variable SMs (Erhardt et al., 2011). Asynchronous multi-dataset analyses can be challenging with this data-driven approach because all the data must be analyzed together. Selecting and labeling the components can also be an arduous task. Spatially constrained ICA (scICA) is an automatic and adaptive approach for estimating the subject-specific features using a priori network templates, hence suitable for multi-dataset analyses. Several algorithms for performing scICA are available in the Group ICA of fMRI Toolbox (GIFT) (<https://trendscenter.org/software/gift/>) (Lin et al., 2010; Du and Fan, 2013). We use the neuromark component templates derived from multiple large-N ($N > 800$) studies combined with the scICA algorithm to estimate features from the individual subjects. These features are then used to classify the subjects based on known medication-class of treatment response. We use an SVM-based algorithm to perform the classification. The algorithm is closely related to prior work in which a single feature (SMs) estimated from the patient data collected at Western University was used (Osuch et al., 2018; Fan et al., 2011). Osuch et al. used known DSM-based BD type-I and MDD patients to create the algorithm, whereas we used the patients' known treatment response (AD vs. MS) to create ours (Osuch et al., 2018).

Our study aims to predict treatment response in patients with mood disorders using resting-state fMRI features. In doing so, we intend to demonstrate the superior efficacy of neuroimaging and data-driven techniques over DSM-based diagnosis. Following are our novel contributions. We use BD and MDD subjects from the Western dataset to create a new SVM-based classification algorithm. We use two independent datasets (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC) and UCLA Consortium for Neuropsychiatric Phenomics LA5c (LA5C)) to validate the trained algorithm (Osuch et al., 2018; Poldrack et al., 2016; Trivedi et al., 2016). We reveal neurophysiological differences in these populations using a stepwise forward feature selection algorithm. This work extends our prior work in several ways (Salman et al., 2021). Previously, we reported results using thresholded SMs as features. Here we use the unthresholded SMs, which lowers the number of false positives and false negatives and results in better sensitivity/specificity. We also fuse the unthresholded SM with functional network connectivity (FNC) to perform multi-feature prediction in Western data and extend the framework in LA5C and EMBARC data. Finally, we also include the prediction of MDD treatment response improvement scores in the EMBARC dataset using the same algorithm. We hope that in the longer term, the algorithm will help predict AD vs. MS response in complex patients with unclear DSM diagnoses. The emphasis on medication-class response potentially provides a clinically useful 'DSM-free' approach to identifying biomarkers of medication-class of response within mood disorders.

2. Methods

2.1. Data

Our medication-class of response predictor model is trained on the resting-state fMRI data collected on MDD and BD patients from Western University. These individuals were followed up over an extended period and categorized based on cumulative knowledge, including medication class response and clinical and research diagnosis. We validated the trained model on two independent datasets: EMBARC and LA5C. These datasets are described below and also summarized in Table 1.

Table 1
Summary of data.

Dataset	Western	LA5C	EMBARC
Population info			
BD	44	49	
MDD	43		545
Nonresponder	14		
Control	39	121	78
Total	140	170	623
Age range	16–27	21–50	18–65
Acquisition parameters			
Scanner type	Siemens	Siemens	GE/Siemens/Phillips
TR	2000 ms	2000 ms	2000 ms
TE	30 ms	30 ms	28 ms
Slices	40	34	
Slice thickness	3 mm	4 mm	3.1 mm
Flip angle	90°	90°	90°
FOV	240 mm	192 mm	205 mm
Matrix size	80 × 80	64 × 64	64 × 64
Scan duration	8 min	304 s	2 × 6 min
Volumes	164	142	178 (–4)

2.1.1. Western data

The University of Western Ontario Research Ethics Board approved the data collection. Written informed consent was obtained from all participants. We divide the data collection into two rounds. The first round of data was collected before 2018 and used in a prior study (Osuch et al., 2018). The subjects were between 16 and 27, with no significant effect of age between groups ($p = 0.1492$). They were divided into four groups: 33 controls, 32 patients with BD type-I, 34 with MDD, and 12 with unknown diagnosis (Osuch et al., 2018). The second round includes data collected at the same site between 2018 and 2021 and the data collected in the first round. The division is because we can use the first round of data in a replication experiment of the prior study. The second round of data includes additional treatment response information to use in our experiments. For the patient group, diagnoses were made using the Structural Clinical Interview for DSM disorders-IV (SCID-IV) or the Diagnostic Interview for Genetic Studies (DIGS). They were confirmed by clinical psychiatric diagnostic assessment. Agreement between SCID-IV/DIGS diagnosis and clinical diagnosis was required for the patients. If there was disagreement between DIGS and clinical diagnosis or if patients had one or more first-degree relatives with mental illness, they were categorized as the "unknown" group. In the second round of data collected until 2021, there were 147 subjects. They were again divided into four groups: 33 controls, 35 patients with BD type-I, 67 with MDD, and 12 with unknown diagnoses.

The medication-class was determined by the clinician using chart review to treat each patient to attain sustained euthymia, lasting at least six months. Medication-class was simplified to either an AD or MS (lithium, lamotrigine, carbamazepine, divalproex sodium) (Osuch et al., 2018). Based on medication class, we divided the 147 subjects into four groups: 33 controls, 47 patients responding to AD, 45 responding to MS, 8 nonresponders, and 14 remitted without medication.

MRI data were acquired at the Lawson Health Research Institute using a 3.0T Siemens Verio MRI scanner and a 32-channel phased-array head coil. The data included gradient-echo, echo-planar imaging (EPI) scans with the following acquisition parameters: repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, 40 axial slices and thickness = 3 mm, with no parallel acceleration, flip angle = 90°, field of view (FOV) = 240 × 240 mm, matrix size = 80 × 80. The length of the resting fMRI scan was approximately 8 min, and 164 brain volumes were collected.

2.1.2. LA5C data

MRI images were collected on two 3.0T Siemens Trio scanners at the Ahmanson-Lovelace Brain Mapping Center (Siemens version syngo MR B15) and the Staglin Center for Cognitive Neuroscience (Siemens version syngo MR B17) at UCLA. 130 healthy individuals from the community and individuals diagnosed with schizophrenia (SZ) (50), BD

(49), and attention deficit hyperactivity disorder (ADHD) (43) participated in the study. The age range of the participants was 21 – 50, and there was a significant effect of age between groups ($p = 0.0025$). fMRI scans were acquired using a T2*-weighted EPI sequence with the following parameters: TR = 2000 ms, TE = 30 ms, slice thickness = 4 mm, 34 slices, oblique slice orientation, flip angle = 90°, matrix 64 × 64, and FOV = 192 mm. Scans covered the whole brain for a total time of 304 s. Previous work can be consulted for additional data descriptions (Poldrack et al., 2016).

2.1.3. EMBARC data

Controls and MDD patients diagnosed using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) at four sites participated in this study. There were 337 subjects in total: 40 controls and 297 patients. The age range of the participants was 18–65. There was no significant age difference among the participants ($p = 0.8015$). Functional imaging was acquired during the resting-state for two scans of 6 min each. The functional image acquisition parameters were: TR = 2000 ms, TE = 28 ms, flip angle = 90°, FOV = 205 mm, slice thickness = 3.1 mm, matrix 64 × 64. Previous work can be consulted for additional data descriptions (Trivedi et al., 2016).

The patients were divided into four groups based on the treatment response on the Clinical Global Improvement (CGI) scale. They were classed as nonresponders if their score was less than 3 or “much improved” (Trivedi et al., 2016). The number of subjects with CGI scores of 1 (no improvement), 2, 3 (much improved), and 4 (completely improved) was 25, 20, 21, and 11 respectively.

2.2. Preprocessing

Data were preprocessed for all three datasets using the Statistical Parametric Mapping (SPM) software (Friston, 2007). The preprocessing steps included rigid body motion correction for head motion, slice-timing correction for the timing difference in slice acquisition, warping into the standard Montreal Neurological Institute (MNI) space using an EPI template, resampling to $3 \times 3 \times 3 \text{ mm}^3$ isotropic voxels, and smoothing using a Gaussian kernel with a full width at half maximum (FWHM) of 6 mm.

We used a state-of-the-art motion correction technique (INRIAlign toolbox in SPM) for motion correction to retain most of the subjects for analysis. We still excluded a small number of subjects from the analysis based on the following quality control criteria. Subjects were discarded if they had more than 3° rotational and 3 mm translational head motion during the scanning period. Only subjects with more than 120 time points were retained. We also ensured that the subjects included for further analysis provided a successful normalization of the whole brain (Fu et al., 2021b).

After preprocessing and quality control, we retained 135 subjects in the second round of Western data, including 13 nonmedicated (remitted without medication), 42 MS responders, 41 CE responders, and 39 controls. As for the LA5C data, 255 subjects were retained: 121 controls, 46 diagnosed with BD, 47 SZ, and 41 ADHD. For this analysis, we used the controls and BD subjects. As for the EMBARC data, we retained 623 subjects (78 controls and 545 diagnosed with MDD).

2.3. Feature extraction

2.3.1. The neuromark template

We used a set of independent component (IC) templates called neuromark (<https://trendscenter.org/data/>) (Du et al., 2020). This reference set comprised 53 labeled and ordered components which were replicated following separate analyses on the control subjects in Human Connectome Project (HCP) and Brain Genomics Superstruct Project (GSP) datasets (Smith et al., 2013; Buckner et al., 2014). The components were divided into seven functional domains: subcortical network (SCN), auditory network (ADN), sensorimotor network (SMN), visual

network (VSN), executive control network (CON), default mode network (DMN), cerebellar network (CBN). Fig. 2 presents a composite view of the neuromark templates. These have been successfully applied in numerous studies and validated as robust spatial priors that provide reliable functional network features across subjects and datasets (Fu et al., 2021c,b,a).

2.3.2. Spatially-constrained ICA

We used the spatially-constrained ICA algorithm available in GIFT (<https://trendscenter.org/software/gift/>) to extract features from pre-processed data of the subjects (Du and Fan, 2013; Lin et al., 2010; Salman et al., 2019). In this fully automated approach, each subject's preprocessed fMRI data were the input, the neuromark fMRI 1.0 templates (available in GIFT and also at <http://trendscenter.org/data>) were used as the reference, and the output included subject-specific SMs and TCs. Furthermore, we estimated the FNC matrix for each subject using the Pearson correlation coefficient between the TCs of 53 components, the dimension of which was 53×53 .

2.4. Classification

2.4.1. Kernel SVM

SVMs are a set of supervised binary classification algorithms which can also be extended for regression and multiclass classification (Vapnik, 1999, 1998). The SVM algorithm incorporates a sample selection mechanism, i.e., only the support vectors affect the decision function. It constructs a maximal margin linear classifier in high-dimensional feature space by mapping the original features via a kernel function. We can define a unique kernel function for applying the SVM algorithm to classify fMRI features such as SMs and TCs. We utilize the similarity measures of subspaces in the commonly used kernel functions (Chang and Lin, 2001).

We initially constructed an SVM kernel matrix in the SM feature space of the subjects and later on constructed kernel matrices on different/multiple feature spaces. The distance metric used in the kernel matrix is the (cosine of the) principal angle between subspaces (PABS) (Björck and Golub, 1973). Let F and G be given subspaces, and $p = \dim(F)$, $q = \dim(G)$, $p \geq q$. Then the principal angles $\theta_k \in [0, \frac{\pi}{2}]$ between F and G are recursively defined for $k = 1, 2, \dots, q$ by

$$\cos \theta_k = \max_{u \in F} \max_{v \in G} u^H v = u_k^H v_k, \|u\|_2 = \|v\|_2 = 1 \quad (1)$$

subject to $u_j^H u = v_j^H v = 0, j = 1, 2, \dots, k-1$

The vectors $(u_1, \dots, u_q)$ and $(v_1, \dots, v_q)$ are principal vectors of the pair of spaces. The spatial ICs estimated using ICA on the subject data constitute a subspace in this SM space. Let F and G correspond to such voxel × component spaces for two subjects, and q be the number of ICs. The (cosine of the) principal angles $\theta_k \in [0, \frac{\pi}{2}]$ between F and G numerically correspond to the ordered singular values of $F^H G$ (Björck and Golub, 1973). Hence, the subspace distance metric can be obtained from the ordered singular values of $F^H G$, i.e.,

$$S(F, G) = \sum_{i=1}^k s_i, \text{ where } s = \text{svd}(F^H G) \quad (2)$$

Finally, the distance matrix S of all subjects is mapped into a high-dimensional feature space K using the sigmoid kernel function, i.e.,

$$K(F, G) = \tanh(\gamma S(F, G)) \quad (3)$$

Fig. 1 presents a flowchart of the classification framework. The (maximally) statistically independent nature of the subject-specific SMs approximates an orthonormal set. It allows us to estimate a Riemannian similarity metric between different subjects and construct a pairwise similarity matrix of all subjects. We then map this similarity matrix into a high-dimensional feature space using a sigmoid kernel function to build the SVM classifier (Björck and Golub, 1973; Fan et al., 2011).

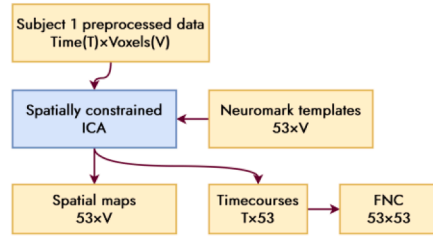
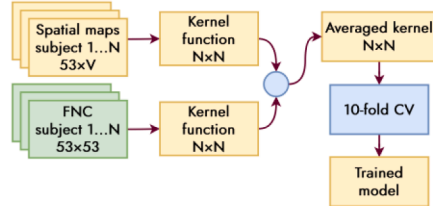
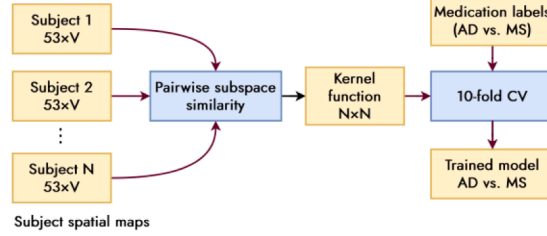
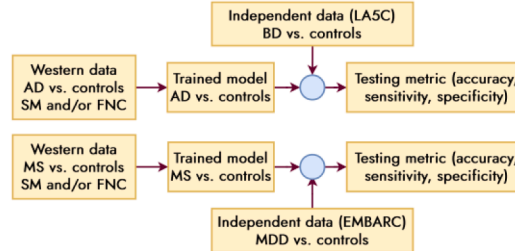
A. Spatially-constrained ICA**C. Multiple kernel SVM****B. Kernel SVM****D. Validation with Independent data**

Fig. 1. Flowchart of our classification scheme. A. Resting-state fMRI data are put through the Neuromark ICA pipeline for feature extraction (spatial maps, time courses, and FNC). B. Classification is performed using kernel SVM algorithm & 10fold cross-validation. Known medication-class of treatment response (mood stabilizers (MS)/antidepressants (AD)) is used as the targets to train the models. C. Experiments are run using spatial maps (SMs), functional network connectivity (FNC), and their combination as features. D. Trained models are tested on independent data.

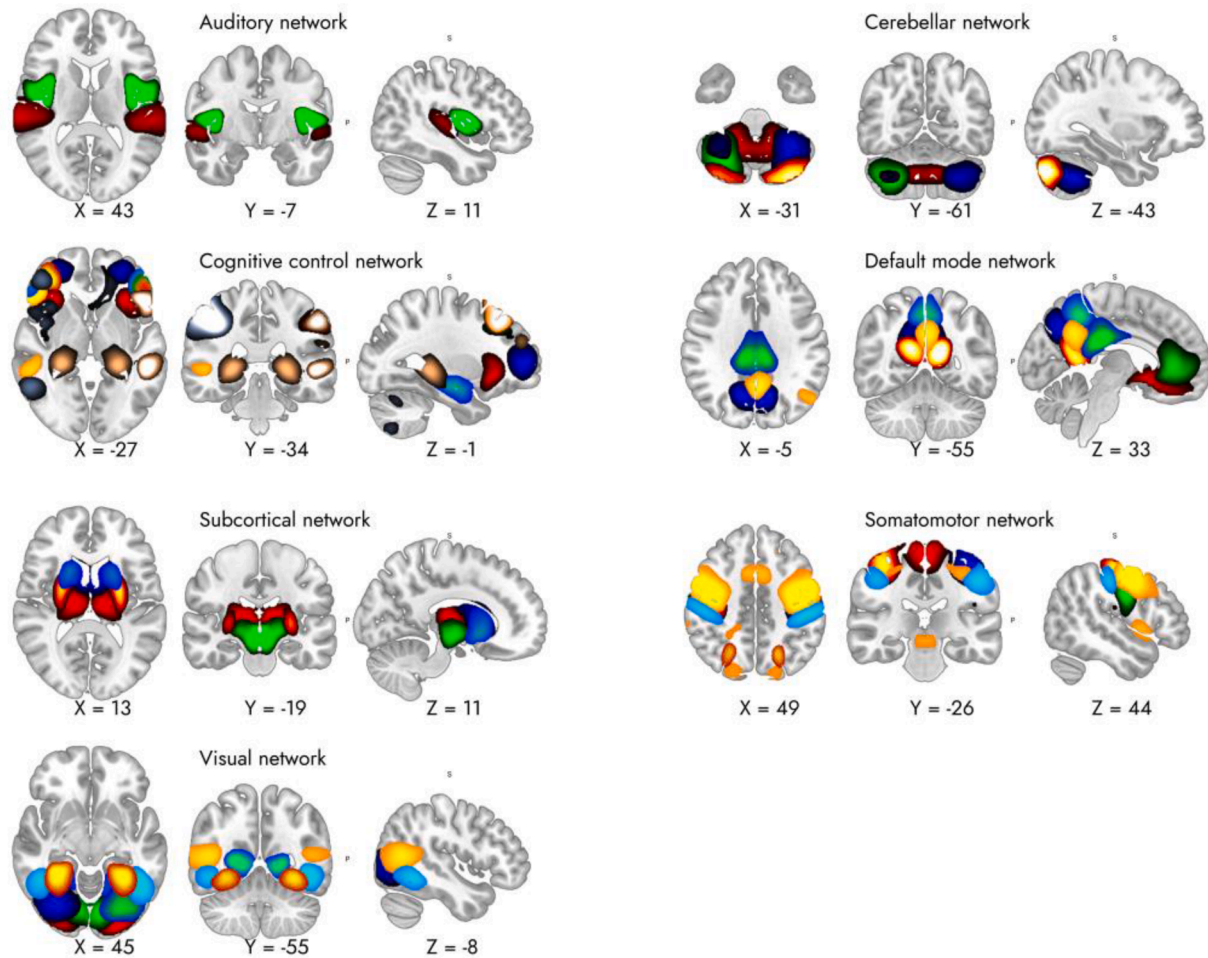


Fig. 2. The neuromark SM templates. These are obtained using group ICA analysis on HCP, GSP controls data, and a greedy algorithm to identify the most replicable components. 53 spatial maps are divided into 7 functional domains. These templates can be used as references to estimate subject spatial maps and time courses from new and unseen data.

2.4.2. Multiple kernel learning

We have features extracted for the same subject from multiple feature spaces (SMs, TCs, FNC, etc.). We can estimate an SVM kernel matrix from each of these underlying subspaces and combine those for even better predictive performance. Here we report results by averaging the kernel matrices as follows. Given two subspaces $(\mathbf{A}_1, \mathbf{A}_2) = (\{\mathbf{a}_{11}, \mathbf{a}_{12}, \dots, \mathbf{a}_{1k}\}, \{\mathbf{a}_{21}, \mathbf{a}_{22}, \dots, \mathbf{a}_{2k}\})$ and $(\mathbf{B}_1, \mathbf{B}_2) = (\{\mathbf{b}_{11}, \mathbf{b}_{12}, \dots, \mathbf{b}_{1k}\}, \{\mathbf{b}_{21}, \mathbf{b}_{22}, \dots, \mathbf{b}_{2k}\})$, where $\mathbf{A}_1, \mathbf{B}_1$ are subspaces for the subjects from one feature space, and $\mathbf{A}_2, \mathbf{B}_2$ are subspaces from another, then we estimate the subspace similarity between the two subjects using the following averaging formula:

$$S_p = \frac{1}{k} \left(\sum_{j=1}^2 \sum_{i=1}^k s_{ji} \right)^{\frac{1}{2}} \quad (4)$$

2.4.3. Forward feature selection

We used the neuromark reference template SMs in the scICA algorithm to extract the subjects' features (SM and TC). The neuromark fMRI 1.0 template includes 53 components divided into seven functional domains. We posit that using a subset (optimal set) of these 53 components may result in a negligible loss in classification score while being less computationally intensive. Therefore we used a stepwise forward selection method to generate such an optimal set (Osuch et al., 2018). The Matlab sequentialfs function implements this algorithm which can be incorporated into a 10-fold cross-validation (CV) scheme. Algorithm 1 contains the pseudocode for implementing this step.

2.5. Main experiment

2.5.1. Prediction of treatment response in western data

We used the neuromark ICA framework to reproduce the result reported in prior work on the first round of data collected at the Western site (Osuch et al., 2018). We used 66 patients' data as the input and their treatment response (32 MS responders and 34 CE responders) classification labels. We performed three experiments with different features from the same subjects: SMs, FNC in the kernel SVM approach, and combined SM+FNC in the multiple kernel learning approach, as outlined in the previous sections. In the next experiment, we used more data from the second collection round at the same site ($N = 83$, 41 MS responders and 42 CE responders) for replication.

The Western data also included two more groups- controls and patients released without medication (nonmedicated). As such, we classified five more pairs of groups: nonmedicated vs. responders (MS or AD), controls vs. responders (MS or AD), and a three-way classification between nonmedicated vs. MS vs. AD. We classified each group using the three different feature spaces (SMs, FNC, and SM+FNC) with 100 repeats and shuffled CV folds at each repetition. This ensures the stability, replicability, and robustness of the reported results. We use stratified cross-validation folds to mitigate the issue of unbalanced samples in responder vs. nonmedicated groups. This ensures that the model has information about all types of samples present in the data at every training step.

2.5.2. Most salient features

Previously, we performed a forward feature selection procedure to reduce the 53 components estimated using neuromark ICA into a smaller set of SMs. We performed this step in every CV step of each repeat

Algorithm 1

Algorithm.

- 1: Start with an empty set $Y_0 = \phi$ and some candidate features
- 2: Select the best next feature $x^+ = \arg \max_{x \in Y_k} [J(Y_k + x)]$
- 3: Update $Y_{k+1} = Y_k + x^+, k = k + 1$
- 4: Return to step 2

experiment, which gave us a collection of the most frequently occurring discriminative SMs. We noted one SM from each functional domain as a salient brain activity pattern for discussion.

2.6. Secondary experiments

2.6.1. Classification of BD vs. controls in LA5C data

We used controls and BD patient data from the LA5C project for a validation experiment. We trained classification models using the controls and patients with MS treatment response labels in Western data. We used these models to predict the controls and patients with BD DSM labels in the LA5C data. Identical to the main experiment, we trained three models with different features from the same subjects (SMs, FNC, and SM+FNC). We repeated the experiments 100 times with shuffled training CV folds. According to Table 1, the LA5C data suffer from an unbalanced sample issue. To mitigate this problem, we used a subject selection step before classification. In this step, we selected an equal number of BD patients and controls from LA5C data.

2.6.2. Classification of MDD vs. controls in EMBARC data

We used controls and MDD patient data from the EMBARC project for another validation experiment. We used the controls and patients with AD treatment response labels in the Western data to train a classification model and to predict the controls and patients with MDD DSM labels in the EMBARC data. Like the previous experiments, we trained three models with different features from the same subjects and repeated them 100 times. According to Table 1 and similar to LA5C, EMBARC data also suffer from an unbalanced sample issue, which we mitigated using a subject selection step.

2.6.3. Classification of MDD improvement scores in EMBARC data

We also classified the patients who responded well (improvement score of 4) and nonresponders (improvement score of 1) in the EMBARC data.

3. Results

3.1. Classification between different groups in western data

Table 2 lists the results of the primary classification experiments based on treatment response labels in Western data. We ran every experiment 100 times with shuffled CV folds for replicability. Below we report the average metrics (accuracy, sensitivity, and specificity) and standard deviations across those runs for each experiment in Table 2. The results are detailed below.

The first row in Table 2 indicates the replication of the result reported by Osuch et al. (2018). We used neuromark-generated features from the SMs of the same 64 subjects to predict treatment response. We obtained a hold-out testing accuracy of 92.8% (sensitivity 0.9, specificity 0.92). In prior work, there were also 12 subjects with unknown diagnoses. Our model obtained 90.9% accuracy in predicting the eventual diagnosis of those "unknown" samples (not shown in Table 2).

When classifying the patients' treatment response based on SM features in the data collected in the second round of acquisition at the same site ($N = 83$), the accuracy was 84.33% (sensitivity 0.87, specificity 0.80). In addition to SM, we also use FNC and a combination of SM and FNC to predict treatment response. We obtained hold-out testing accuracy of 85.5% using FNC (sensitivity 0.91, specificity 0.79), and 86% using SM+FNC (sensitivity 0.9, specificity 0.81).

When classifying nonmedicated subjects from MS responders using SM features, we obtained 93.3% accuracy (sensitivity 0.99, specificity 0.71, $N = 54$). Using FNC features, the accuracy was 97.1% (sensitivity 1.0, specificity 0.86), and using SM+FNC it was also similar (97.1% accuracy, sensitivity 1.0, specificity 0.87). The same classification approach with AD responders resulted in 92.4% accuracy (sensitivity 0.99, specificity 0.68, $N = 53$). Using FNC features, the accuracy was

Table 2

Main classification experiment results.

Groups <i>Osuch et al. (2018)</i>	Features	Training N	Testing scores (hold-out data)		
			Accuracy	Sensitivity	Specificity
MS-AD	SM	64	92.8±1.9	0.90	0.92
Replication using Neuromark ICA					
MS-AD	SM	83	84.3±3.3	0.87±0.03	0.80±0.05
	FNC		85.5±2.9	0.91±0.01	0.79±0.05
	SM+FNC		86.0±2.6	0.90±0.01	0.81±0.05
MS-nonmedicated	SM	54	93.3±2.9	0.99±0.01	0.71±0.12
	FNC		97.1±1.3	1.00	0.86±0.05
	SM+FNC		97.1±1.2	1.00	0.87±0.05
AD-nonmedicated	SM	53	92.4±2.3	0.99±0.01	0.68±0.11
	FNC		97.3±0.9	1.00	0.88±0.04
	SM+FNC		97.5±1.1	1.00	0.88±0.04
MS-AD-nonmedicated	SM	95	80.2±2.9	0.83±0.02	0.81±0.07
	FNC		88.2±2.5	0.87±0.01	0.89±0.05
	SM+FNC		87.3±2.6	0.86±0.01	0.87±0.05

97.3% (sensitivity 0.99, specificity 0.88), and using SM+FNC it was 97.5% (sensitivity 1.0, specificity 0.88). In a threeway classification among MS, AD responders, and the nonmedicated subjects ($N = 95$), we obtained 80.2% accuracy (sensitivity 0.83, specificity 0.81).

3.2. Classification in independent data

3.2.1. Classification of BD in LA5C data

Table 3 lists the results of the secondary classification experiments on LA5C and EMBARC data. We performed a validation experiment with the controls and BD population of the LA5C dataset. Using only the SM features, the model achieved a hold-out testing accuracy of 79.8% (sensitivity 0.50, specificity 0.90). Using the FNC features, the hold-out testing accuracy was 82.6% (sensitivity 0.63, specificity 0.89), and using the combination SM+FNC features, it was 82.5% (sensitivity 0.63, specificity 0.89). The accuracy was 85.9% for BD-diagnosed patients in Western data (sensitivity 0.85, specificity 0.85). Using the FNC features, the hold-out testing accuracy was 85.7% (sensitivity 0.91, specificity 0.79), and using the combination SM+FNC features, it was 85.2% (sensitivity 0.9, specificity 0.79).

3.2.2. Classification of MDD in EMBARC data

We performed another validation experiment with the controls and MDD population of the EMBARC dataset. In this experiment, the model achieved a hold-out testing accuracy of 89% (sensitivity 0.88, specificity 0.89). Using the FNC features, the hold-out testing accuracy was 87.6% (sensitivity 0.0, specificity 1.0), and using the combination SM+FNC features, it was 85.8% (sensitivity 0.9, specificity 0.81). The accuracy was 85.2% for MDD-diagnosed patients in Western data (sensitivity 0.84, specificity 0.85). Using the FNC features, the hold-out testing accuracy was 84.5% (sensitivity 0.9, specificity 0.78), and using the combination SM+FNC features, it was 84.4% (sensitivity 0.9, specificity

Table 3

Secondary (classification) experiment results.

Groups	Features	Training N	Testing scores (hold-out data)		
BD-control (LA5C)	SM	166	79.8±1.8	0.50±0.06	0.90±0.02
	FNC		82.6±0.8	0.63±0.01	0.89±0.01
	SM+FNC		82.5±0.8	0.63±0.02	0.89±0.01
BD-control (Western)	SM	81	85.9±3.4	0.85±0.04	0.85±0.04
	FNC		85.7±2.6	0.91±0.01	0.79±0.04
	SM+FNC		85.2±2.5	0.90±0.02	0.79±0.05
MDD-control (EMBARC)	SM	77	89.0±3.7	0.88±0.02	0.89±0.06
	FNC		87.6	0	1
	SM+FNC		85.8±3.4	0.90±0.01	0.81±0.06
MDD-control (Western)	SM	80	85.2±3.0	0.84±0.04	0.85±0.03
	FNC		84.5±2.6	0.90±0.02	0.78±0.05
	SM+FNC		84.4±2.5	0.90±0.02	0.78±0.05

Classification Based on MDD Patient Improvement Scores in EMBARC data.

0.78).

The algorithm was able to separate patients with an improvement score of 1 (no improvement) and 4 (completely improved) with an accuracy of 69.44% (sensitivity 0.96, specificity 0.09).

3.3. Most salient features

Fig. 3 shows multi-planar views of the most salient neuromark templates. The corresponding subject-level features of these templates were the best-performing predictors of treatment response in Western data. The Automated Anatomical Labeling (AAL) labels for these templates are the following: superior temporal gyrus from the ADN network (volume 21 in the neuromark template), cerebellum from the CBN network (13), inferior parietal lobule from the CON network (68), precuneus from the DMN network (32), caudate from the SCN network (69), postcentral gyrus from the SMN network (3), and calcarine gyrus from the VSN network (16).

4. Discussion

In this work, we report our algorithm's hold-out and testing performance based on features extracted from resting-state fMRI data. We also report the independent validation performance by testing the algorithm on the LA5C and EMBARC data.

Several areas of agreement exist between the most salient (neuromark) spatial maps and prior work (Osuch et al., 2018). Five ICs were identified as the most salient features for classification. These included bilateral inferior parietal lobule, posterior DMN regions, anterior cingulate cortex, a combination of caudate, thalamus, and parahippocampal gyrus, and lastly, the insular region. In our experiments, we also found, among others, the right inferior parietal lobule, precuneus (in the posterior DMN region), and caudate regions to be salient.

We replicated these results using the scICA approach based on the neuromark template (Du et al., 2020). The prior work used the group-informationguided ICA (GIG-ICA) framework, meaning the group-level components were estimated from the same dataset on which the classification was performed. The advantage of the scICA framework over GIG-ICA is that it is an adaptive approach using a priori network templates, suitable for multidataset analyses (Salman et al., 2019).

We had the DSM diagnosis and the medication-class of treatment response (AD or MS) data available for the patients in the Western data. We used the latter (medication-class) as the hold-out/testing dataset labels for developing the model. We also ran separate classification experiments on the independent datasets (LA5C and EMBARC). However, the DSM-based diagnosis labels were the classification target values in those datasets. In doing so, we demonstrate that the model trained on treatment response data can also predict the DSM diagnosis, although

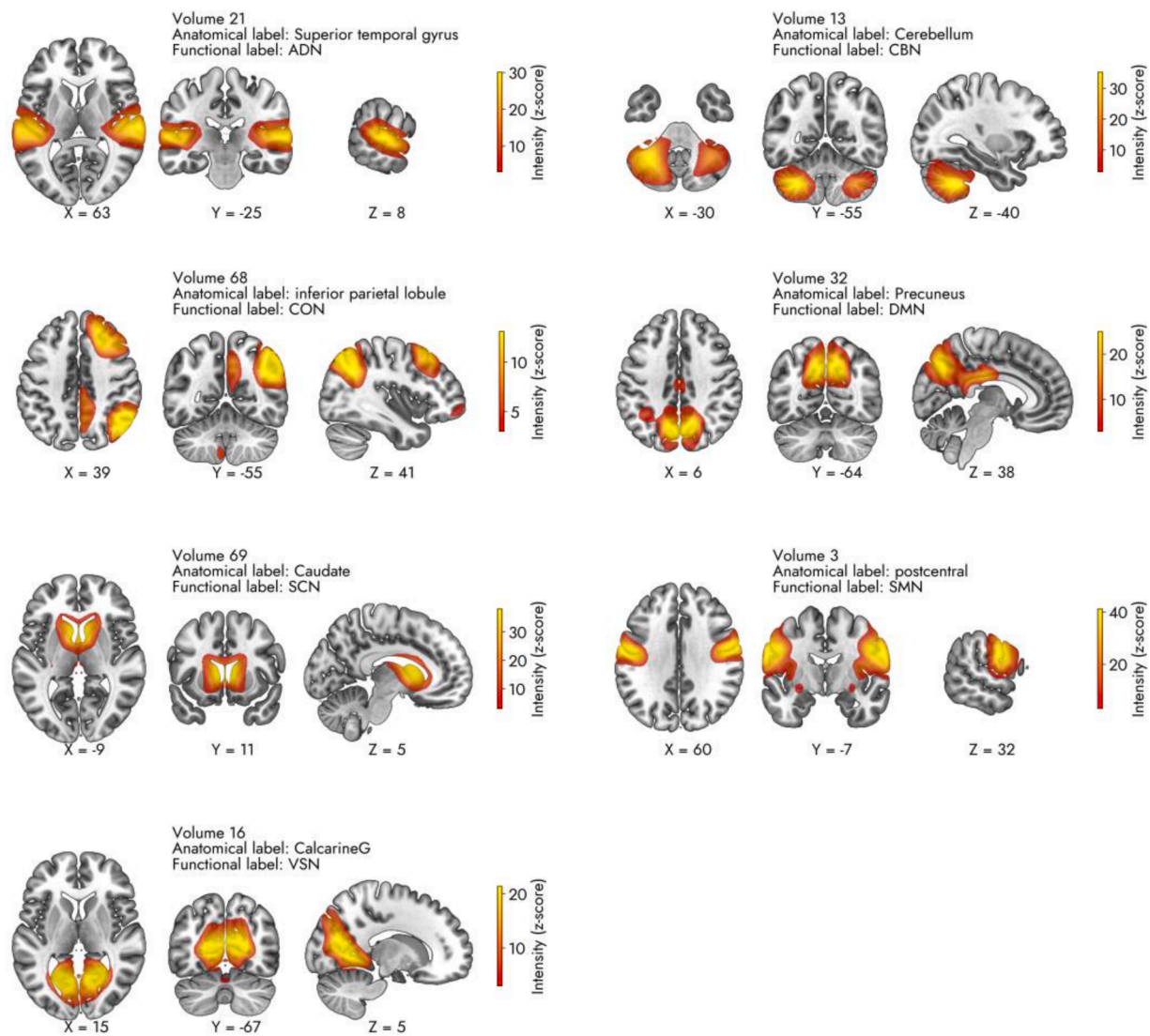


Fig. 3. Best component(s) for treatment response prediction in each functional domain across different experiments.

slightly less accurate.

The other strengths of our approach include using a template derived from higher model order group ICA analysis resulting in more granular SMs. Also, we leveraged the TCs (or FNC) separately and combined them with the SMs in a multiple kernel learning framework. We report FNC-based classification results in each alternate row of Table 2. The FNC outperformed the SM in classifying these cohorts in all experiments.

Another versatile feature of this model is the ability to predict the MDD improvement scores with reasonable accuracy. The low specificity of most of these experiments indicates comparatively high false positives than false negatives when detecting the patients' improvement scores, which is more desirable clinically. The other strength is the ability to predict the unknown samples from the first round of acquisition in Western data. In prior work, there were 12 subjects with unknown diagnoses (Osuch et al., 2018), and our model obtained 90.9% accuracy in predicting their eventual diagnosis. It indicates the robust efficacy and utility of the model.

5. Limitations

We will discuss some of the limitations of the study next. There is an age difference across the datasets. However, we use the analysis of variance (ANOVA) test to show no significant effect of age on the

diagnosis or treatment response variables in two of the three datasets used. Harmonization techniques such as ComBat can mitigate the effect of varying acquisition parameters at various sites (Johnson et al., 2007; Fortin et al., 2017, 2018; Bostami et al., 2022). The treatment response information is available from Western data only; the other sites provide DSM diagnosis. Efforts should be directed at collecting more data with treatment response information included in the future to validate similar results. The LA5C dataset contained individuals with diagnoses of ADHD and SZ. Although these were not included in the analysis, these comparator groups can be informative as symptoms and treatment response overlap with mood disorders in future studies.

We generated a kernel function consisting of multiple modalities using Eq. (4). A weighted approach or multiple kernel learning method can significantly improve this process (Tanabe et al., 2008; Gonen and Alpaydin, 2011). The Riemannian distance measure is most useful when orthonormal basis vectors span the subspaces. However, the FNC feature space consists of scalar values only. Moreover, one of the assumptions of spatial ICA algorithms is that the components are maximally statistically independent. It is valid for the SMs but not necessarily for the TCs. Orthonormal features from the TCs may allow the kernel method to perform better.

Some of the experiments conducted suffer from an unbalanced sample issue. In such a case, conveying the method's efficacy is

impossible using only the accuracy metric. We have addressed this issue by sampling balanced subjects, using stratified cross-validation folds, and reporting more meaningful metrics such as sensitivity and specificity.

We may treat the SMs and FNC as separate modalities in machine learning. However, both are estimated from a single neuroimaging modality (fMRI). Future multi-modal analysis may rely on including data from structural magnetic resonance imaging (sMRI), diffusion tensor imaging (DTI), and other neuroimaging modalities.

6. Conclusion

The goal of our study was to predict treatment response in patients with mood disorders using resting-state fMRI features. In doing so, we demonstrated the superior efficacy of neuroimaging and data-driven techniques. The algorithm will help predict AD vs. MS response in complex patients with unclear DSM diagnoses. The emphasis on medication-class response potentially provides a clinically useful 'DSM-free' approach to identifying biomarkers of medication-class of response within mood disorders.

Declaration of Competing Interest

The authors report no biomedical financial interests or potential conflicts of interest.

Acknowledgement

This work was supported by the National Institutes of Health grants 5R41MH122201 and R01MH118695 (to Calhoun V.D.) and National Science Foundation grant 2112455 (to Calhoun V.D.).

References

- American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th edition ed. American Psychiatric Association. <https://doi.org/10.1176/appi.books.9780890425596>.
- Björck, A., Golub, G.H., 1973. Numerical methods for computing angles between linear subspaces. *Math. Comput.* 27, 579–594. <https://doi.org/10.2307/2005662> arXiv: 2005662.
- Bostami, B., Hillary, F.G., van der Horn, H.J., van der Naalt, J., Calhoun, V.D., Vergara, V.M., 2022. A decentralized ComBat algorithm and applications to functional network connectivity. *Front. Neurol.* 13, 826734. <https://doi.org/10.3389/fneur.2022.826734>.
- Bowden, C.L., 2005. A different depression: clinical distinctions between bipolar and unipolar depression. *J. Affect. Disord.* 84, 117–125. [https://doi.org/10.1016/S0165-0327\(03\)00194-0](https://doi.org/10.1016/S0165-0327(03)00194-0).
- Buckner, R.L., Roffman, J.L., Smoller, J.W., 2014. Brain Genomics Superstruct Project (GSP). doi:10.7910/DVN/25833.
- Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. <https://doi.org/10.1002/hbm.1048>.
- Chang, C.C., Lin, C.J., 2001. Training nu-support vector classifiers: theory and algorithms. *Neural Comput.* 13, 2119–2147. <https://doi.org/10.1162/089976601750399335>.
- de Almeida, J.R.C., Phillips, M.L., 2013. Distinguishing between unipolar depression and bipolar depression: current and future clinical and neuroimaging perspectives. *Biol. Psychiatry* 73, 111–118. <https://doi.org/10.1016/j.biopsych.2012.06.010>.
- Du, Y., Fan, Y., 2013. Group information guided ICA for fMRI data analysis. *Neuroimage* 69, 157–197. <https://doi.org/10.1016/j.neuroimage.2012.11.008>.
- Du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., Salman, M., Abrol, A., Rahaman, M.A., Chen, J., Hong, L.E., Kochunov, P., Osuch, E.A., Calhoun, V.D., 2020. NeuroMark: an automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *NeuroImage Clin.*, 102375. <https://doi.org/10.1016/j.nicl.2020.102375>.
- Erhardt, E.B., Rachakonda, S., Bedrick, E.J., Allen, E.A., Adali, T., Calhoun, V.D., 2011. Comparison of multi-subject ICA methods for analysis of fMRI data. *Hum. Brain Mapp.* 32, 2075–2095. <https://doi.org/10.1002/hbm.21170>.
- Fan, Y., Liu, Y., Wu, H., Hao, Y., Liu, H., Liu, Z., Jiang, T., 2011. Discriminant analysis of functional connectivity patterns on Grassmann manifold. *Neuroimage* 56, 2058–2067. <https://doi.org/10.1016/j.neuroimage.2011.03.051>.
- Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* 167, 104–120. <https://doi.org/10.1016/j.neuroimage.2017.11.024>.
- Fortin, J.P., Parker, D., Tun, C.B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T., 2017. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>.
- Friston, K.J., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, 1st ed. Elsevier/Academic Press, Amsterdam; Boston.
- Fu, Z., Iraj, A., Sui, J., Calhoun, V.D., 2021a. Whole-brain functional network connectivity abnormalities in affective and non-affective early phase psychosis. *Front. Neurosci.* 15, 682110. <https://doi.org/10.3389/fnins.2021.682110>.
- Fu, Z., Iraj, A., Turner, J.A., Sui, J., Miller, R., Pearlson, G.D., Calhoun, V.D., 2021b. Dynamic state with covarying brain activity-connectivity: on the pathophysiology of schizophrenia. *Neuroimage* 224, 117385. <https://doi.org/10.1016/j.neuroimage.2020.117385>.
- Fu, Z., Sui, J., Turner, J.A., Du, Y., Assaf, M., Pearlson, G.D., Calhoun, V.D., 2021c. Dynamic functional network reconfiguration underlying the pathophysiology of schizophrenia and autism spectrum disorder. *Hum. Brain Mapp.* 42, 80–94. <https://doi.org/10.1002/hbm.25205>.
- Gönen, M., Alpaydm, E., 2011. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* 12, 2211–2268.
- Han, S., Cui, Q., Wang, X., Li, L., Li, D., He, Z., Guo, X., Fan, Y.S., Guo, J., Sheng, W., Lu, F., Chen, H., 2020. Resting state functional network switching rate is differently altered in bipolar disorder and major depressive disorder. *Hum. Brain Mapp.* 41, 3295–3304. <https://doi.org/10.1002/hbm.25017>.
- He, Z., Sheng, W., Lu, F., Long, Z., Han, S., Pang, Y., Chen, Y., Luo, W., Yu, Y., Nan, X., Cui, Q., Chen, H., 2019. Altered resting-state cerebral blood flow and functional connectivity of striatum in bipolar disorder and major depressive disorder. *Progr. Neuro-Psychopharmacol. Biol. Psychiatry* 90, 177–185. <https://doi.org/10.1016/j.pnpbp.2018.11.009>.
- Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. <https://doi.org/10.1093/biostatistics/kxj037>.
- Judd, L.L., Akiskal, H.S., Schettler, P.J., Endicott, J., Maser, J., Solomon, D.A., Leon, A.C., Rice, J.A., Keller, M.B., 2002. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch. Gen. Psychiatry* 59, 530. <https://doi.org/10.1001/archpsyc.59.6.530>.
- Lin, Q.H., Liu, J., Zheng, Y.R., Liang, H., Calhoun, V.D., 2010. Semiblind spatial ICA of fMRI using spatial constraints. *Hum. Brain Mapp.* 31, 1076–1088. <https://doi.org/10.1002/hbm.20919>.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188. [https://doi.org/10.1002/\(SICI\)1097-0193\(1998\)6:3<160::AID-HBMS5>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBMS5>3.0.CO;2-1).
- Osuch, E., Gao, S., Wammes, M., Théberge, J., Williamson, P., Neufeld, R.J., Du, Y., Sui, J., Calhoun, V.D., 2018. Complexity in mood disorder diagnosis: fMRI connectivity networks predicted medication-class of response in complex patients. *Acta Psychiatr. Scand.* 138, 472–482. <https://doi.org/10.1111/acps.12945>.
- Perlis, R.H., Brown, E., Baker, R.W., Nierenberg, A.A., 2006. Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *AJP* 163, 225–231. <https://doi.org/10.1176/appi.appi.163.2.225>.
- Poldrack, R., Congdon, E., Triplett, W., Gorgolewski, K., Karlsgodt, K., Mumford, J., Sabb, F., Freimer, N., London, E., Cannon, T., Bilder, R., 2016. A phenotype-wide examination of neural and cognitive function. *Sci. Data* 3, 160110. <https://doi.org/10.1038/sdata.2016.110>.
- Rai, S., Griffiths, K.R., Breukelaar, I.A., Barreiros, A.R., Chen, W., Boyce, P., Hazell, P., Foster, S.L., Malhi, G.S., Harris, A.W.F., Korgaonkar, M.S., 2021. Default-mode and fronto-parietal network connectivity during rest distinguishes asymptomatic patients with bipolar disorder and major depressive disorder. *Transl. Psychiatry* 11, 1–8. <https://doi.org/10.1038/s41398-021-01660-9>.
- Salman, M.S., Du, Y., Lin, D., Fu, Z., Fedorov, A., Damaraju, E., Sui, J., Chen, J., Mayer, A., Posse, S., Mathalon, D., Ford, J.M., Van Erp, T., Calhoun, V.D., 2019. Group ICA for identifying biomarkers in schizophrenia: 'Adaptive' networks via spatially constrained ICA show more sensitivity to group differences than spatio-temporal regression. *NeuroImage Clin.*, 101747. <https://doi.org/10.1016/j.nicl.2019.101747>.
- Salman, M.S., Verner, E., Bockholt, H.J., Fu, Z., Calhoun, V.D., 2021. Machine learning predicts treatment response in bipolar & major depression disorders. In: 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), pp. 1–6. <https://doi.org/10.1109/BIBE52308.2021.9635339>.
- Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D.C., Glasser, M.F., 2013. Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168. <https://doi.org/10.1016/j.neuroimage.2013.05.039>.
- Tanabe, H., Bao, H., Nguyen, C.H., Kawasaki, S., 2008. Simple but effective methods for combining kernels in computational biology. In: 2008 IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies. IEEE, Ho Chi Minh City, Vietnam, pp. 71–78. <https://doi.org/10.1109/RIVF.2008.4586335>.
- Trivedi, M.H., McGrath, P.J., Fava, M., Parsey, R.V., Kurian, B.T., Phillips, M.L., Quendo, M.A., Bruder, G., Pizzagalli, D., Toups, M., Cooper, C., Adams, P., Weyandt, S., Morris, D.W., Grannemann, B.D., Ogden, R.T., Buckner, R., McInnis, M., Kraemer, H.C., Petkova, E., Carmody, T.J., Weissman, M.M., 2016. Establishing moderators and biosignatures of antidepressant response in clinical care

- (EMBARC): rationale and design. J. Psychiatr. Res. 78, 11–23. <https://doi.org/10.1016/j.jpsychires.2016.03.001>.
- Vapnik, V., 1998. The support vector method of function estimation. In: Suykens, J.A.K., Vandewalle, J. (Eds.), Nonlinear Modeling: Advanced Black-Box Techniques. Springer US, Boston, MA, pp. 55–85. https://doi.org/10.1007/978-1-4615-5703-6_3.
- Vapnik, V., 1999. An overview of statistical learning theory. IEEE Trans. Neural Netw. 10, 988–999. <https://doi.org/10.1109/72.788640>.