



# Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data



Xing Meng<sup>a,1</sup>, Rongtao Jiang<sup>a,1</sup>, Dongdong Lin<sup>b</sup>, Juan Bustillo<sup>c</sup>, Thomas Jones<sup>c</sup>, Jiayu Chen<sup>b</sup>, Qingbao Yu<sup>b</sup>, Yuhui Du<sup>b</sup>, Yu Zhang<sup>a</sup>, Tianzi Jiang<sup>a,e</sup>, Jing Sui<sup>a,b,e,\*</sup>, Vince D. Calhoun<sup>b,c,d,\*</sup>

<sup>a</sup> Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>b</sup> The Mind Research Network and Lovelace Biomedical and Environmental Research Institute, Albuquerque, NM 87106, USA

<sup>c</sup> Dept. of Psychiatry and Neuroscience, University of New Mexico, Albuquerque, NM 87131, USA

<sup>d</sup> Dept. of Electronic and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA

<sup>e</sup> CAS Center for Excellence in Brain Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

## ARTICLE INFO

### Article history:

Received 28 December 2015

Revised 13 April 2016

Accepted 7 May 2016

Available online 10 May 2016

### Keywords:

Individualized prediction

Multimodal

MATRICES Consensus Cognitive Battery (MCCB)

Schizophrenia

MRI

Neuromarker

## ABSTRACT

Neuroimaging techniques have greatly enhanced the understanding of neurodiversity (human brain variation across individuals) in both health and disease. The ultimate goal of using brain imaging biomarkers is to perform individualized predictions. Here we proposed a generalized framework that can predict explicit values of the targeted measures by taking advantage of joint information from multiple modalities. This framework also enables whole brain voxel-wise searching by combining multivariate techniques such as ReliefF, clustering, correlation-based feature selection and multiple regression models, which is more flexible and can achieve better prediction performance than alternative atlas-based methods. For 50 healthy controls and 47 schizophrenia patients, three kinds of features derived from resting-state fMRI (fALFF), sMRI (gray matter) and DTI (fractional anisotropy) were extracted and fed into a regression model, achieving high prediction for both cognitive scores (MCCB composite  $r = 0.7033$ , MCCB social cognition  $r = 0.7084$ ) and symptomatic scores (positive and negative syndrome scale [PANSS] positive  $r = 0.7785$ , PANSS negative  $r = 0.7804$ ). Moreover, the brain areas likely responsible for cognitive deficits of schizophrenia, including middle temporal gyrus, dorsolateral prefrontal cortex, striatum, cuneus and cerebellum, were located with different weights, as well as regions predicting PANSS symptoms, including thalamus, striatum and inferior parietal lobule, pinpointing the potential neuromarkers. Finally, compared to a single modality, multimodal combination achieves higher prediction accuracy and enables individualized prediction on multiple clinical measures. There is more work to be done, but the current results highlight the potential utility of multimodal brain imaging biomarkers to eventually inform clinical decision-making.

© 2016 Elsevier Inc. All rights reserved.

## Introduction

Predictive data mining has become a powerful tool for researchers and clinical practitioners in medicine and neuroscience. It involves building and applying theory and methods that allow for effective creation, evaluation and selection of prediction models (Dumpry et al., 2010; Oakes et al., 2007; Simpson et al., 2014). Neuroimaging has greatly enhanced our understanding of the human brain and its variation across individuals (neurodiversity) in health and disease (Edwards et al., 2011; Oto et al., 2011; Wang et al., 2011b; Yan et al., 2013). However, most criteria that used to assess severity/prognosis of brain disorders are still primarily based on clinical judgment; thus, the subjective factors of the doctors cannot be avoided. In addition, the overlapped

cognitive or behavior performance of several mental disorders emphasizes the inadequacy of a diagnosis based purely on symptoms or behaviors alone and highlights the need for objective imaging neuromarkers that can assist the timely diagnosis or treatment. Therefore, using brain-imaging data to identify the neuroanatomical basis of cognitive impairment or symptom changes in brain diseases is an important research topic, since it is helpful for better understanding of the pathophysiology underlying the illness.

Predictive data mining has become popular in neuroimaging studies in a recent decade, especially for mental disorders research. Recent studies have started to use machine-learning techniques to detect neuroimaging patterns that may predict cognitive or behavioral performance (Aharoni et al., 2013; Eichele et al., 2008; Shen et al., 2014). And initial brain measures have indicated compelling potential to predict health-related outcomes (Chen et al., 2007; Clark et al., 2014; Gabrieli et al., 2015; Shaffer et al., 2013; Technow et al., 2014; Tsang et al., 2009; Vittengl et al., 2014; Willette et al., 2014). However, most predictive studies to date have only related variation in baseline brain

\* Corresponding author at: Institute of Automation, Chinese Academy of Sciences, Beijing, China.

E-mail address: [jing.sui@nlpr.ia.ac.cn](mailto:jing.sui@nlpr.ia.ac.cn) (J. Sui).

<sup>1</sup> Xing Meng and Rongtao Jiang contributed equally as the first authors.

measures to variation in subsequent outcomes, which could be described more as post-diction or correlation rather than prediction (Gabrieli et al., 2015; Hutton et al., 2009). There are, however, a few examples of studies that are indeed “predicting” specific values of the target measures. For example, Wan et al. has proposed an elegant regression model called CORNLIN (Wan et al., 2014) that employs a sparse Bayesian learning algorithm to predict multiple cognitive scores based on 98 structural MRI regions of interests (ROIs) for Alzheimer's disease patients. The polynomial model used in CORNLIN can detect either a nonlinear or linear relationship between brain structure and cognitive decline. Stonnington et al. adopted relevance vector regression (Stonnington et al., 2010), a sparse kernel method formulated in a Bayesian framework, to predict four sets of cognitive scores using MRI voxel based morphometry measures. Wang et al. (Wang et al., 2011a) and Zhang et al. (Zhang et al., 2012a) employed multi-task learning strategies for selecting biomarkers that could predict multiple clinical scores, e.g., using  $\ell_2$ -norm coupled with  $\ell_1$ -norm in regression models or multi-task feature selection coupled with a support vector machine (SVM). Integrating and testing various types of baseline data available, Ye et al. applied sparse logistic regression with stability selection to ADNI (Alzheimer's Disease Neuroimaging Initiative) data for robust feature selection (Ye et al., 2012), successfully predicted the conversion from MCI to probable AD and identified a small subset of bio-signatures.

Nevertheless, many works were limited to use of a single modality, without using cross information from multiple modalities or were focused on subgroup discrimination (“classification”) instead of the particular values of the measures (Gabrieli et al., 2015; Leergaard et al., 2012). In addition, quite often the input feature dimensionality is restricted, e.g. based on ROI features derived from AAL (Anatomical Automatic Labeling) templates, but without performing whole brain voxel-wise searching because of the computational load. Hence, we are motivated to propose a generalized model that is able to predict explicit values of the targeted measures by whole brain voxel-wise searching via cutting-edge machine learning algorithms (Stokes and Visweswaran, 2012; Sun, 2007) and by combining joint information from multiple imaging modalities. Based on this motivation, we could use purely data-driven techniques to estimate the cognitive scores symptom changes or to classify disease subcategories using MRI measures, providing a potentially biologically valid framework for understanding mental illnesses or for even predicting the progression of the disease. This is consistent with the scope of the research domain criteria (RDoC) project proposed by NIMH, which aims to develop new ways of classifying and clarifying the underlying causes of mental disorders based on dimensions of observable behavior and neurobiological measures (Cuthbert and Workgroup, 2014).

Here, we aim to achieve two main goals: 1) Design a generalized prediction framework that can be widely employed in future studies. 2) Predict target clinical measures quantitatively and identify relevant neuromarkers that could be biologically meaningful on cognitive decline/functional deficits in schizophrenia (SZ). For real human brain imaging data, we applied the framework to 47 SZ patients and 50 healthy controls (HCs) to estimate their cognitive scores, measured by the “measurement and treatment research to improve cognition in schizophrenia” (MATRICS) Consensus Cognitive Battery (MCCB) (August et al., 2011) and the patients' symptom scores, measured by the positive and negative syndrome scale (PANSS) (Kay et al., 1987), using features extracted from three modalities: fractional amplitude of low frequency fluctuations (fALFF) from resting-state functional MRI, segmented gray matter (GM) from structural MRI and fractional anisotropy (FA) from diffusion MRI.

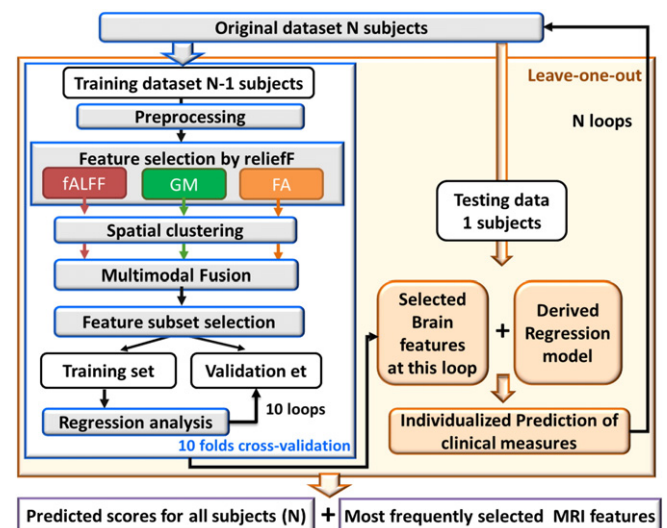
Note that we used the same data set as used for (Sui et al., 2015), however, different than the previous study that investigated the multimodal co-varied networks which were not only significantly correlated with cognitive composite scores, but also showed significant differences between schizophrenia and healthy controls; this study aims to explore an approach to multimodal imaging data that enables detection of

neuroimaging patterns as well as the building of a prediction model of target measures, such as cognitive scores or symptom scores. The methods used are quite distinct, the former paper adopted multi-set canonical correlation analysis, a blind source separation approach applied to group level of subjects; while the current work utilized several machine-learning techniques, including feature selection, clustering, regression and cross-validation. Results suggest the proposed prediction framework shows great promise to produce precise and individualized estimates on multiple clinical measures. Hence it's possible to use this tool to estimate the treatment outcome for an individual patient, e.g., remitter or non-remitter, based on observable imaging measures, similar to what did in (van Waarde et al., 2015), which could enable the clinician to realistically make the personalized medical decision before treating a patient. Therefore, this project will be one of few attempts to meet clinical challenges of making early intervention possible based on fundamental neuroscience.

## Materials and methods

### Theory development

The working flowchart of proposed prediction model is shown in Fig. 1, which uses three MRI measures as an example; but this surely can be extended to other measures or modalities. The basic idea and justification of each step can be described as follows: First, considering the individuality of the predicted measures (cognition, symptoms), we remove features with very few variability across all training subjects by relative standard deviation thresholding, then based on advanced feature selection method ReliefF (Stokes and Visweswaran, 2012), we are able to select a set of most relevant features to the predicted measures. Then data-driven spatial clustering was performed to obtain fewer ROI-based features by averaging of the clusters, of which one cluster may consist several parts of the brain regions parcellated by atlas-based template, such AAL template. Note that using brain regions segmented by atlas as features is an alternative in our prediction, whereas, our proposed method is a purely data-driven feature extraction via whole brain voxel-wise searching, and achieves a higher prediction accuracy compared to AAL atlas-based methods, see results and the supplementary file (Supplementary Table S4). After this, the ROI-based features of different modalities were combined to take advantage of the multimodal complementary information, and further refined by correlation-based feature selection



**Fig. 1.** The working flowchart of the proposed prediction framework. Fig. 1 shows working flowchart of the proposed prediction model, in which preprocessing, feature selection (ReliefF), spatial clustering, feature subset selection and regression models are employed. Three MRI measures (fALFF, GM, FA) are combined together to realize the individualized prediction via linear or nonlinear regression analysis.

(CFS) (Tripoliti et al., 2010), producing a feature subset before regression. This step is necessary; since spatial clustering and averaging generate ROIs that compared to voxel-wise features indicate decreased relevance to target measure and may contain redundant information. Finally, by cutting-edge multiple regression models, a combination of joint brain feature subsets is formed to best predict the targeted measures. Note that all above mentioned feature selection algorithms (Relieff, CFS) and regression models can be found in WEKA (Frank et al., 2004), a powerful data mining software in Java with a collection of machine learning algorithms. We will also release our prediction toolbox in future.

To avoid introducing bias into the prediction, rigorous nested cross-validation was performed. First, we left one subject for testing and used the other 96 subjects for training; then in the training layer, the 96 subjects went through a 10-fold cross-validation in the regression analysis after initial feature selection and the multimodal combination as shown in Fig. 1. This loop was repeated 97 times to test all the subjects. Each time the results produced the predicted score for the left-out (test) subject, as well as the regression equation and the identified brain regions in each modality. Then, we calculate the accuracy of the resulting prediction and identify the frequently occurring brain regions. Finally, in order to get a single regression equation for the whole MCCB prediction, we also performed a 10-fold prediction using all subjects to illustrate the identified features and regression model, as shown in Fig. 3. Next we describe the details for each step.

#### Feature preprocessing

Generally speaking, specific features with only minor variation across all subjects will not be useful for prediction. By setting an appropriate threshold, we eliminate features with low standard deviation (std), which also helps reduce the dimensionality. This step is similar to variance-based thresholding that has been popularly used in DNA methylation pre-processing for dimension reduction (Wilhelm-Benartzi et al., 2013). The std. threshold is set as  $th = 0.5 \times \frac{\text{std}(\text{target})}{\text{mean}(\text{target})} \times \text{mean}(\text{all features})$  where “target” denotes the measure to be predicted. The necessity of this step is given in supplementary file, see Table S1.

#### Feature selection by regression Relieff

After obtaining representative measures from each modality, we will estimate the relevance of each voxel-wise feature to the target measure by Relieff (Stokes and Visweswaran, 2012) to derive subset of features. For continuous response values here, regression Relieff is adapted. When processing, assume the input feature matrix  $\mathbf{X}$  is in dimension of subjects by voxels ( $N \times L$ ), Relieff randomly selects one subject  $R_i$  and searches for its  $k$  nearest neighbor subjects  $I_j$  ( $j = 1, 2, \dots, k$ ) based on  $L1$  distance. For one feature (voxel/attribute)  $A$ , if it is desirable for prediction, then a bigger distance between the selected subject  $R_i$  and its nearest neighbors  $I_j$  should correspond to farther values on their predicted measures. For short, Relieff introduces a kind of probability that the predicted values of two participants are different. This probability can be modeled with the relative distance between the predicted values of two subjects, and then are used to estimate the importance (weight) of each feature for the regression (Robnik-Šikonja and Kononenko, 1997).

Assume  $N_{dc}$  measures two nearest subjects (e.g.,  $R_i$  and  $I_j$ ) have different predictions,  $N_{d[A]}$  determines two nearest participants have different values for the voxel/feature  $A$ ,  $P_{dc \& d[A]}$  measures two nearest participants have different predictions and different values for voxel feature  $A$ .  $N_{dc}$ ,  $N_{d[A]}$  and  $P_{dc \& d[A]}$  were updated  $m$  times for all voxels. Finally, the weight for each feature  $A$  is defined as:

$$W[A] = \frac{N_{dc \& d[A]}}{N_{dc}} - \frac{N_{d[A]} - N_{dc \& d[A]}}{m - N_{dc}} \quad (1)$$

While, the detailed updates for  $N_{dc}$ ,  $N_{d[A]}$  and  $N_{dc \& d[A]}$ , as well as the Pseudo code of RRelieff can be found in (Robnik-Šikonja and Kononenko, 2003).

Following regression Relieff, every feature is assigned with a weight value statistically accounts for its relevance to the response values (prediction). By determining a certain output number according to the scree test (Mori et al., 2000) (see more details in the supplementary file Supplementary Fig. S1), we can identify the most significant features in the dataset and use them as the representatives of the original voxels.

#### Spatial clustering

The voxels selected via Relieff were in general distributed sparsely across the whole brain. Here we employ a 26-connected ( $3 \times 3 \times 3 - 1$ ) neighborhood strategy for image dilation (Bazin et al., 2011), a classic way to find connected voxels in 3D image (Cheng et al., 2009). In our case, the voxels selected out by Relieff is 1, and the other voxels are 0. The 26-connected neighborhood voxels are neighbors to every voxel that touches one of their faces, edges, or corners. These pixels are connected along either one, two, or all three of the primary axes. These new generated voxel clusters can be projected into their corresponding areas on a brain map, enabling a useful visualization of the results. Next, these clusters are treated as the elementary features, and the mean voxel values of each cluster represent the ROI-based feature value.

#### Feature fusion

Features selected from the above procedures are derived from a single modality, reflecting a view of either brain structure or function (Sui et al., 2012). By combining features of different modalities together in a joint analysis, potentially important variations or relationship that may only be partially detected by single modality (Sui et al., 2013; Sui et al., 2014) could be revealed, and complementary information may be obtained for better prediction (Arbabshirani et al., 2016; Calhoun and Sui, 2016). At this step, in order to take advantage of joint information, all ROI-based brain features of 3 modalities were concatenated horizontally, to build a matrix in the dimension of  $N_{\text{subjects}}$  by ( $N_{\text{feat\_fMRI}} + N_{\text{feat\_sMRI}} + N_{\text{feat\_dMRI}}$ ) and then input to the feature subset selection.

#### Feature subset evaluation

Although voxel-wise features selected by Relieff are important, after spatial clustering and extract mean of the clusters, some features may have little influence on the predicted attribute. To further remove the redundancy, correlation-based feature selection (CFS) is employed (Tripoliti et al., 2010). In CFS, the worth of each subset of features is evaluated by considering the individual predictive ability of each feature along with the degree of redundancy between them (Bielza and Larranaga, 2014; Gudmundsson et al., 2012). Subsets of features that are highly correlated with the predicted measure while having low inter-correlation are preferred (Kohavi and John, 1997). Such a feature selection can help reduce the size of the resulting knowledge structures based on multimodal features and further yield a concise and refined data set that significantly improves the prediction performance.

#### Regression analysis

Next, a final regression analysis that quantitatively captures correlations between target measures and selected brain features was implemented. If we have  $l$ ,  $p$ ,  $q$  number of features derived from fALFF, GM and FA, respectively, then the practical predicted values (not true target



value) can be written as:

$$\text{Predicted value} = \beta + \theta_1 \cdot fALFF_1 + \theta_2 \cdot fALFF_2 + \dots + \theta_l \cdot fALFF_l + \eta_1 \cdot GM_1 + \eta_2 \cdot GM_2 + \dots + \eta_p \cdot GM_p + \lambda_1 \cdot FA_1 + \lambda_2 \cdot FA_2 + \dots + \lambda_q \cdot FA_q \quad (2)$$

To make a performance comparison between linear and nonlinear regressions, we adopt three models: multiple linear regression, pace regression (linear) and sequential minimal optimization regression (SMOreg) (non-linear), all of which can be called and implemented through WEKA too. The latter two models are briefly introduced as follows.

#### *PACE regression*

(Projection Adjustment by Contribution Estimation) has striking advantages over existing techniques for fitting linear models with a strong emphasis on dimensionality determination problem. Compared to classical ordinary least squares estimators that may fail to detect redundancy in a larger feature set, pace regression improves it by evaluating the effect of each variable and using a clustering analysis to improve the statistical basis for estimating their contribution to the overall regressions. Besides, pace regression adopts nonnegative-measure-based minimum distance method for solving the minority cluster problem. It can yield better prediction models with reduced model dimensionality (Meshkin et al., 2009), especially for high dimensional data. Detailed information can be found in (Wang, 2000).

#### *Sequential minimal optimization regression*

(SMOreg) is an iterative algorithm (Schölkopf and Smola, 2002) for solving the regression problem using support vector machines (SVM). SMOreg uses the same principles as the SVM for classification, except for applying a nonlinear polynomial kernel  $k(x_i, x_j) = (x_i, x_j)^p$  with  $p = 2$  to transfer features into higher dimensional space, to make it possible to perform the linear separation. The goal of SMOreg is to estimate a function that is as “close” as possible to target values and as “flat” as possible for good generalization. More details together with a pseudo-code can be found in (Li and Jiang, 2006). As an extension of the SMO algorithm proposed by (Platt, 1999) for SVM classifier design, SMOreg overcomes the issue of an important source of confusion and inefficiency caused by SMO. It globally normalizes all attributes by default (Mohammadi et al., 2013), with an excellent performance on handling big samples.

#### *Human brain data*

##### *Participants*

Participants included a total of 47 patients with a SCID-P based DSM-IV-TR diagnosis of schizophrenia and 50 age-matched HCs (Table 1). The study was IRB approved at all participating institutions: the University of New Mexico Hospital and the Albuquerque Veterans Administration Medical Center. Participants were free of any central neurological disorder or significant head trauma and have no current diagnosis of substance abuse (>6 months before enrollment, excluding nicotine). Patients were clinically stable with no recent medication change (>4 months prior to study enrollment) across the data collection period. HCs and their first-degree relatives had no history of psychiatric disorder.

##### *Clinical measures*

To measure the current cognitive functioning, a trained rater administered the MCCB within 1 week of imaging. A composite score was calculated via the MCCB scoring program, which is an equal weighting of seven MCCB domain scores and has been recognized as the optimal primary outcome measure (Burton et al., 2013). Compared with HCs, SZs

**Table 1**

Demographics and the MCCB scores of the subjects.

Measure	HC	SZ	P*	R*
Number	50	47		
Age	36.7 ± 12.6	35.3 ± 12.6	0.6	0.04
Gender	20F/30M	6F/41M	0.01	0.17
MCCB Composite	49.8 ± 10.5	31.3 ± 15.7	1.3E-09	1
Speed of processing	51.9 ± 9.0	35.3 ± 13.7	1.5E-09	0.91
Working memory	46.8 ± 11.4	37.1 ± 14.5	5.3E-04	0.83
Attention/Vigilance	48.3 ± 9.9	36.0 ± 15.1	1.4E-05	0.86
Visual learning	49.3 ± 9.3	36.6 ± 12.6	1.5E-07	0.79
Verbal learning	47.4 ± 8.9	38.0 ± 8.6	8.4E-07	0.8
Social cognition	50.8 ± 11.1	40.5 ± 13.0	8.3E-05	0.65
Reasoning/Problem solving	54.2 ± 9.9	46.1 ± 11.7	5.1E-04	0.64
PANSS Positive	NA	15.4 ± 5.9		−0.10
PANSS Negative	NA	15.1 ± 5.4		−0.48

MCCB = MATRICS Consensus Cognitive Battery; PANSS = positive and negative syndrome scale. Olanzapine equivalent = olanzapine total (standardized current dose of antipsychotic medication).

P\* denotes the significance value of two sample t-tests performed between controls and schizophrenia patients for MCCB scores. R\* is the correlation value between MCCB composite score and other measures.

achieved significantly lower scores in composite and all domains. Meanwhile, the PANSS was conducted to quantify symptom scores of patients. The negative PANSS scores and the MCCB composite were anti-correlated ( $R = -0.48$ ,  $p = 0.0008$ ) as expected. No significant correlation was found between MCCB composite and medication dose in SZ.

#### *Imaging parameters*

All subjects were scanned by fMRI, sMRI and diffusion magnetic resonance imaging (dMRI), which were collected on a 3-Tesla Siemens Trio scanner with a 12-channel radio frequency coil.

##### *fMRI*

Resting-state scans were a minimum of 5 min in duration (152 volumes). Subjects were instructed to keep their eyes open during the scan and stare passively at a presented fixation cross, as this facilitates network delineation compared to eyes-closed conditions and helps ensure that subjects are awake. Data were collected with single-shot full k-space echo-planar imaging with ramp sampling correction using the inter-commissural line (anterior commissure/posterior commissure) as a reference (TR = 2 s, TE = 29 ms, matrix size =  $64 \times 64$ , flip angle =  $75^\circ$ , slice thickness = 3.5 mm, slice gap = 1.05 mm, field of view (FOV) = 240 mm, matrix size =  $64 \times 64$ , voxel size =  $3.75 \times 3.75 \times 4.55 \text{ mm}^3$ ).

##### *sMRI*

A multi-echo MPRAGE sequence was used with the following parameters: TR/TE/TI = 2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900 ms, flip angle =  $7^\circ$ , FOV =  $256 \times 256 \text{ mm}$ , slice thickness = 176 mm, matrix size =  $256 \times 256 \times 176$ , voxel size =  $1 \times 1 \times 1 \text{ mm}$ , pixel bandwidth = 650 Hz, total scan time = 6 min.

##### *dMRI*

Data were collected along the anterior commissure/posterior commissure line throughout the whole brain with the following parameters: FOV =  $256 \times 256 \text{ mm}$ , slice thickness = 2 mm, number of excitations = 1, TE = 84 ms, TR = 9000 ms. A multiple channel radio frequency coil was used with generalized auto calibrating partially parallel acquisition ( $\times 2$ ), 35 gradient directions,  $b = 800 \text{ s/mm}^2$  and 5 measurements with  $b = 0$ .

##### *fMRI preprocessing*

The SPM8 software package (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8>) was employed to perform fMRI preprocessing. Slice

timing was performed with the middle slice as the reference frame. Images were realigned using INRI align (Freire et al., 2002). The fMRI data were then despiked to mitigate the impact of outliers and spatially normalized into the standard Montreal Neurological Institute space (Friston et al., 1995) with slightly up-sampled to  $3 \times 3 \times 3 \text{ mm}^3$ . We further regressed out six motion parameters, white matter (WM) and CSF in de-noising, and the mean frame wise displacements showed no significant group difference (mean of root of mean square frame-to-frame head motions assuming 50 mm head radius (Allen et al., 2011)); HC:  $0.224 \pm 0.12 \text{ mm}$ , SZ:  $0.227 \pm 0.12 \text{ mm}$ ,  $p = 0.91$ ). Finally, data were and slightly subsampled to  $3 \times 3 \times 3 \text{ mm}^3$  and spatially smoothed with a Gaussian kernel with full-width half maximum (FWHM) of  $8 \times 8 \times 8 \text{ mm}^3$ . For the rest-fMRI, we extracted the voxel-wise fALFF to generate a map for each subject (Calhoun and Allen, 2012).

#### dMRI preprocessing

Data were preprocessed by the FMRIB Software Library ([www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). All images were registered to the first  $b = 0$  image by the FMRIB linear image registration tool. The preprocessing consisted of the following steps: (a) quality check, any gradient directions with excessive motion or vibration artifacts were identified and removed; (b) motion and eddy current correction; (c) correction of gradient directions for any image rotation done during the previous motion correction step; and (d) calculation of diffusion tensor and scalar measures such as FA, which were then resampled to  $3 \times 3 \times 3 \text{ mm}^3$  and smoothed by a Gaussian kernel with FWHM of  $8 \times 8 \times 8 \text{ mm}^3$  (Sui et al., 2011).

#### sMRI preprocessing

Data were preprocessed using the SPM8 software package to segment the brain into WM, GM, and cerebral spinal fluid with unmodulated normalized parameters via the unified segmentation method (Ashburner and Friston, 2005). According to (Radua et al., 2014), the use of unmodulated GM maps is one of the optimal settings in voxel-based morphometry analysis. Then GM images were then resampled to  $3 \times 3 \times 3 \text{ mm}^3$  and smoothed by a Gaussian kernel with FWHM of 8 mm (White et al., 2001). Subject outliers were detected using a spatial Pearson correlation with the template image to ensure that all subjects were properly segmented (Segall et al., 2009).

#### Normalization

After preprocessing, the 3D brain images of each subject were reshaped into a one-dimensional vector and stacked, forming a matrix ( $N_{\text{subj}} \times N_{\text{voxel}}$ ) for each of the three modalities. These three matrices were then normalized to have the same average sum-of-squares (computed across all subjects and all voxels/locus for each modality) to ensure that all modalities had comparable ranges. Since SZ and HC were not perfectly gender matched in the current study, before prediction, the gender was regressed out to remove its potential influence on differences between groups, even though the correlation between gender and the MCCB composite score is not significant ( $r = 0.17$ ,  $p > 0.05$ ).

#### Individualized prediction

Matrices derived from the above processing could then be treated as the original feature sets, and the corresponding cognitive/symptom scores are treated as the targeted measures; together, they serve as input to the proposed framework. Take fMRI for example, thresholding by standard deviation will prune out almost half of the total number of features, and about thousands of features will be reserved after ReliefF. During spatial clustering, with voxels organized into larger clusters, the number of features will be reduced to dozens. Finally, the features from three modalities are combined into a concatenated feature matrix that

is further refined via subset feature selection, resulting in 5–15 brain regions and achieving the final prediction equation by regression analysis. In the experiment, we predicted the MCCB composite score and two domain scores (social cognition and verbal learning) with rigorous cross-validation, then evaluated PANSS positive and negative scores to verify the validity and reliability of the proposed framework.

In order to ensure the validity of the detected features, we performed an unbiased prediction flow based on nested cross-validation (10 fold + leave one out) as shown in Fig. 1, the correlations and root mean squared errors between the predicted measures and the true values are calculated in both training and testing loops, as shown in supplementary Table S3. For each modality, we could get a set of ROIs that contribute to regression and occur frequently in all loops. Note that the regression models are different at each loop; in order to get one regression model for the whole prediction, we also performed a prediction of MCCB composite using all subjects with 10 fold cross-validation. In this prediction, although in ReliefF all the training subjects were used, the selected voxel-wise features were not directly input into the pace regression, instead, these voxels were spatially clustered and refined again. Such a flow actually reduced the degree of overfitting in the training. We want to check if the frequently occurred ROI features from unbiased tests are in consistence with what we obtained from the above processing. And as we tested, the results obtained from such a working flow, as shown in Fig. 3, are quite similar to what we got from the unbiased test, see Results section.

On the other hand, the proposed method enables data-driven, voxel-wise feature searching. Indeed, using brain regions segmented by an atlas, e.g., AAL, as features can be used as an alternative. For fair comparison, we also added an alternative prediction of MCCB by combining AAL features from GM and fALFF with LASSO (least absolute shrinkage and selection operator), which is a penalized linear regression model popularly used for feature subset selection and making relevant predictions (Tibshirani, 1996; Tibshirani, 2011). We did not use DTI data since the AAL parcellation is based on gray matter structure, which is not suitable for FA maps representing the white matter tracts. The correlation between predicted values with the ground truth, the root mean squared prediction error (RMSE) as well as the normalized root mean squared prediction error (NRMSE) were calculated for each method using the nested cross-validation (10 fold + leave one out), as listed in Table 3. It clearly indicates that the proposed method outperformed the others in all perspectives.

## Results

#### Impact of feature selection options and method comparison

Table 2 compares the prediction accuracy on MCCB composite values, PANSS positive and negative scores using three regression models based on the proposed feature selection framework (the first 3 columns, ReliefF + CFS + regression). Results showed pace regression outperforms multivariate linear regression and nonlinear regression (SMOreg) in building a prediction model for the MCCB composite, achieving the highest correlation at  $r = 0.7033$ , while for PANSS prediction, the nonlinear regression model yields the best results with  $r > 0.7785$  for both positive and negative scores. Since no single regression approach is clearly superior in all cases during prediction process (Langley and Simon, 1995), we will incorporate all three regression models into the prediction toolbox as options for users in our future work. In addition, to demonstrate the necessity of each feature selection steps, we also added the MCCB/PANSS prediction performance based only on SVOREG without employing different feature selection steps (the 4th column) and without CFS step (ReliefF + SMOreg, the 5th column). It's clear that lacking any of the steps result in lower correlation with ground truth. Specifically, if using only SMOreg without feature selection, tens of thousands of voxels were used to build a regression, resulting in markedly worse estimation. Similarly, without CFS, an

**Table 2**

Performance comparison with different feature selection steps and regression models.

Correlation	Proposed + linear reg	Proposed + pace reg	Proposed + SMOreg	SMOreg only	Relieff + SMOreg	SVR only
Modality	MCCB composite					
<b>GM</b>	0.3401	0.5678	0.3152	0.1707	0.3297	0.1689
<b>ALFF</b>	0.5213	0.5677	0.5578	0.2411	0.2242	0.2051
<b>DTI</b>	0.2813	0.3373	0.3627	−0.0559	0.0992	−0.1390
<b>3-WAY</b>	0.6366	<b>0.7033</b>	0.6887	0.2807	0.3579	0.1313
<b>Modality</b>	<b>PANSS positive</b>					
<b>GM</b>	0.6807	0.6972	0.6800	−0.2395	0.6380	−0.2553
<b>ALFF</b>	0.616	0.6322	0.6327	0.1071	0.5139	0.0312
<b>2-WAY</b>	0.674	0.6668	<b>0.7785</b>	−0.1361	0.6701	−0.2662
<b>Modality</b>	<b>PANSS negative</b>					
<b>GM</b>	0.6143	0.7003	0.7334	−0.3664	0.1733	−0.3594
<b>ALFF</b>	0.5561	0.6072	0.7517	−0.0598	0.1257	−0.0552
<b>2-WAY</b>	0.6573	0.7073	<b>0.7804</b>	−0.3975	0.2217	−0.3800

Table 2 demonstrates the necessity of the multi-stage feature selection. The first 3 columns show the performance using the proposed feature selection steps (std thresholding + Relieff + CFS) with different regression models. In addition, we also added the MCCB/PANSS prediction performance using only SVM-type nonlinear regression models (SMOreg) or support vector regression models (SVR) without different feature selection steps. If only employ SVM-based regression, there are tens of thousands of voxels used to build a regression within each modality, resulting in markedly lower correlations between the estimated scores and true MCCB/PANSS scores. We have added these results to the supplementary file to show the necessity of each step in the workflow.

essential step to exclude redundant ROI features, correlations between the estimated scores and true MCCB/PANSS scores decrease substantially too. Besides, supplementary Table S1 shows that including a simple step, standard deviation thresholding, the prediction performance will be improved greatly compared to not use it, especially when the predicted target shows adequate individuality; since most of the stationary features across subjects will be removed at first. All above results demonstrated the validity of the proposed method for multiple predicting targets.

Table 3 indicates the performance of MCCB prediction using the atlas based features instead of voxel-wise searching, as well as comparison with alternative prediction methods. The fALFF and GM data were segmented into 116 ROIs for each subject based on automated anatomical labeling (AAL) (Natsopoulos et al., 1998), and then input into 3 regression models similarly as we did in the proposed method. We also added an alternative prediction by combining AAL features with LASSO (least absolute shrinkage and selection operator), which is a penalized linear regression model popularly used for feature subset selection and making relevant predictions (Tibshirani, 1996; Tibshirani, 2011). We did not use DTI data here since the AAL parcellation is based on gray matter structure, which is not suitable for FA maps representing the white matter tracts. The correlation between predicted values with the ground truth, the root mean squared prediction error (RMSE) as well as the normalized root mean squared prediction error (NRMSE) ([https://en.wikipedia.org/wiki/Root-mean-square\\_deviation](https://en.wikipedia.org/wiki/Root-mean-square_deviation)) were calculated for each method using the nested cross-validation (10 fold + leave one out). As listed, the proposed method outperformed all other alternatives, suggesting that the proposed voxel-wise searching enables more flexibility and higher precision than atlas-based features in prediction. Finally, supplementary Table S2 recorded the training and testing results for MCCB composite prediction on both correlations and RMSE mentioned above, which indicated that there is no apparent overfitting in our processing.

**Table 3**

Comparison between AAL-based feature selection with the proposed method.

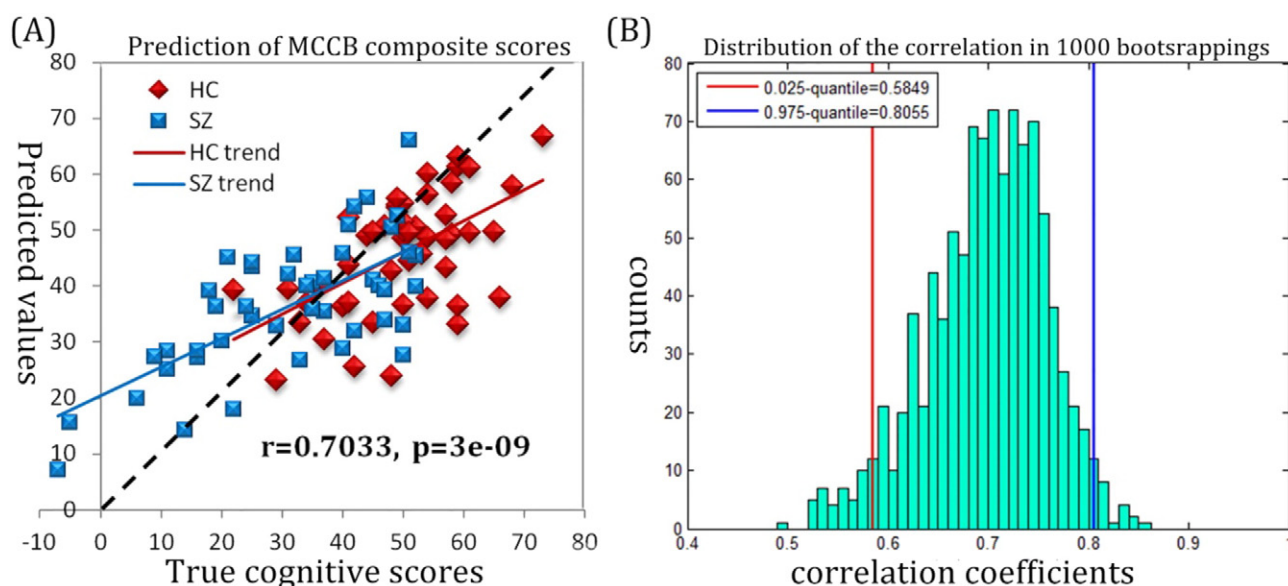
Methods	AAL + pace regression	AAL + Relieff + CFS pace regression	AAL + CFS pace regression	AAL + LASSO	Proposed method
Modality	Correlation between predicted MCCB values and the ground truth				
ALFF	0.16	0.44	0.40	0.34	0.57
GM	0.13	0.26	0.16	0.35	0.57
Modality	Root mean squared prediction error (RMSE)				
ALFF	16.7	14.5	14.8	15.0	13.1
GM	16.5	15.5	16.5	15.8	13.4
Modality	Normalized root mean square error (NRMSE)				
ALFF	0.21	0.18	0.19	0.19	0.16
GM	0.21	0.19	0.21	0.20	0.17

### Prediction of cognitive scores

Fig. 2 demonstrates the predicted cognitive scores for 47 SZs and 50 HCs with respect to their true values based on the flowchart shown in Fig. 1, with a correlation coefficient of  $r = 0.7033$ . The true MATRICS scores are HC:  $50 \pm 11$  and SZ:  $31 \pm 16$ , while RMSE = 11, low enough to separate most of the two groups. As shown in Fig. 2(A), the predicted score accuracy (measured by  $1 - |1 - \text{predicted value}/\text{true value}|$ ) is  $>75\%$  for 2/3 of the subjects and  $>90\%$  for 1/3 of the subjects. Fig. 2(B) demonstrates the bootstrapping test result, which was performed 1000 times on all subjects, and the bootstrapped confidence interval was [0.5849, 0.8055] at 95% (Qin et al., 2015). Connections with a bootstrap ration over 3.0 were considered to be significantly correlated with the cognitive scores.

Table 4 lists the anatomical information and the occurring frequency of the identified brain ROI features for MCCB composite prediction in 97 loops. Brodmann areas (BAs) of the fALFF and GM features and the WM tracts (from the John Hopkins Atlas) overlapped with FA maps are listed according to their occurring frequency from high to low. The higher frequency, the more robustness and importance of the brain regions contribute to the prediction. Clearly, the left prefrontal gyrus (BA 6, 8), middle temporal gyrus (MTG, BA 21 22), visual cortex (BA 18,19), inferior parietal lobule (BA 7,40) and cerebellum are identified as the significant fMRI or sMRI regions contributing to cognitive composite prediction, while for dMRI, parts of superior longitudinal fasciculus, uncinate fasciculus (UF) and inferior longitudinal fasciculus (ILF) seem to play more important roles.

Fig. 3 illustrates the identified multimodal ROI features by using all subjects to build a single regression equation. The identified clusters in each modality as well as their weights resulted from the final pace regression model are demonstrated, including six fALFF clusters, four GM clusters and five FA clusters. The anatomical information of these



**Fig. 2.** Prediction results based on three multimodal features for MCCB MATRICS. Fig. 2: (A) prediction plot based on the proposed framework, cognitive scores (the MCCB composite) for 47 SZs and 50 HCs were plotted in the x-axis (HC: red dots, SZ: blue dots); the Y-axis represents the predicted values. (B) In the reliability test, 1000 bootstrap resamples were performed to estimate the distribution and 95% confidence intervals of correlation coefficients between cognitive score and neuroimaging.

15 ROIs are listed in supplementary Table S3, which shows considerable overlaps with the ROIs listed in Table 4 from unbiased testing. More importantly, prediction accuracy is greatly improved by combining three MRI measures (i.e.,  $r = 0.7033$  using 3-way fusion compared to  $r = 0.5677$  using only fALFF,  $r = 0.5678$  using only GM and  $r = 0.3373$  using only FA, respectively, as shown in Table 2), which is

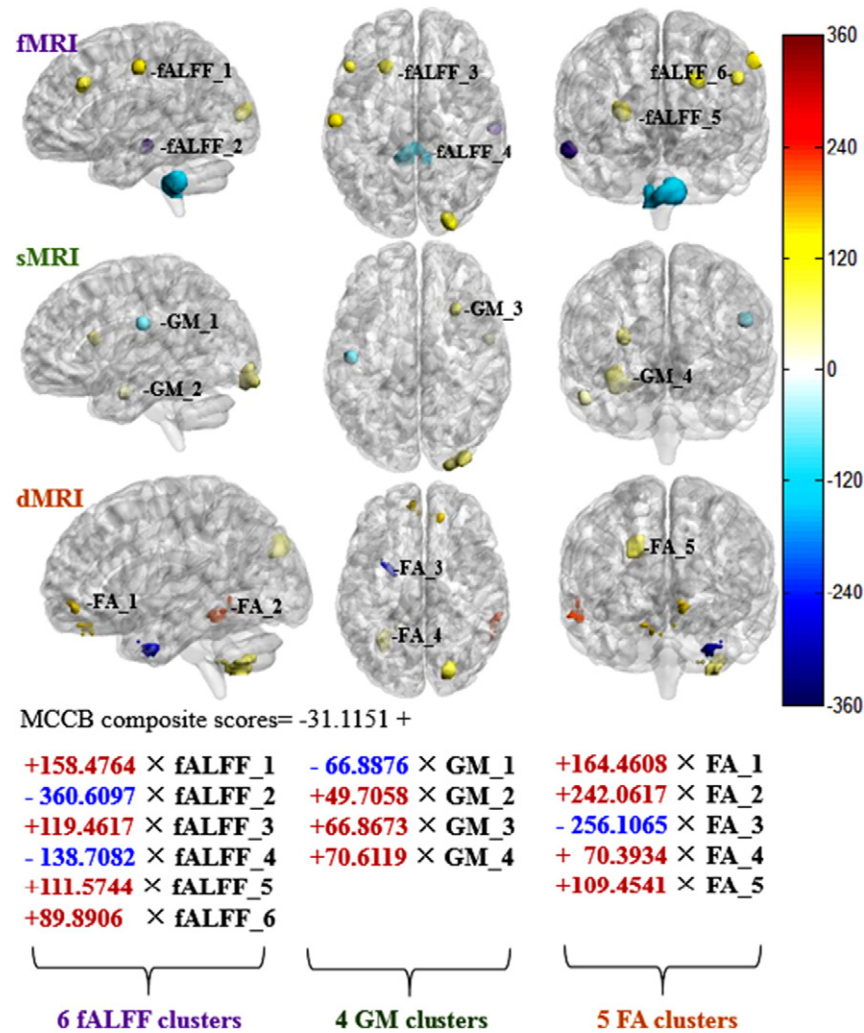
consistent with the known benefits of multimodal fusion (Calhoun and Adali, 2009; Sui et al., 2012). Besides the MCCB composite, there are 7 specific cognitive domains. The proposed method can actually predict all domains at an accuracy  $>0.6$ , here we chose one domain, social cognition, as an example. Its final correlation with ground truth is  $r = 0.7084$  based on an fMRI–sMRI fusion ( $r = 0.4966$  using only

**Table 4**

Anatomical information and the frequency of the identified brain ROI features occurring in 97 loops for MCCB composite prediction ( $r = 0.7033$ ).

Area	Brodmann area	Vol (cc) L/R	Frequency L/R (x, y, z)
<b>fMRI</b>			
Superior temporal gyrus	22	0.0/0.8	NA/0.95 (59, −20, 2)
Middle temporal gyrus	21	0.0/0.3	NA/0.94 (59, −20, −3)
Middle frontal gyrus	8	0.7/0.0	0.74 (−46, 26, 31)/NA
Superior frontal gyrus	6	0.1/0.0	0.50 (−21, 23, 44)/NA
Cerebellum		1.5/0.1	0.44 (−11, −45, −33)/0.21 (13, −45, −33)
Parahippocampal gyrus	36	0.4/0.0	0.42 (−27, −35, −12)/NA
Cuneus	18, 19	0.0/0.3	NA/0.42 (26, −91, 28)
Anterior cingulate	25	0.3/0.0	0.25 (−3, 3, −4)/NA
Precentral gyrus	6	0.1/0.0	0.15 (−58, −11, 42)/NA
<b>sMRI</b>			
Inferior parietal lobule	7, 40	0.2/1.0	0.35 (−38, −34, 42)/0.77 (45, −57, 44)
Cuneus/Precuneus	7, 18, 31	1.2/0.6	0.63 (−15, −70, 30)/0.63 (27, −97, −1)
Middle frontal gyrus	6, 8	0.0/0.6	NA/0.6 (36, 23, 31)
Posterior cingulate		0.1/0.0	0.50 (−3, −53, 8)/NA
Cerebellum		6.2/4.6	0.50 (−3, −54, 0)/0.50 (6, −48, 2)
Lingual gyrus/fusiform	18, 19	0.6/0.4	0.50 (−18, −69, −10)/0.21 (36, −70, 17)
Postcentral gyrus	3	0.8/0.0	0.48 (−50, −17, 30)/NA
Parahippocampal gyrus	35	0.1/0.3	0.07 (−35, −11, −18)/0.27 (30, −23, −19)
Middle Temporal gyrus	21	0.0/0.7	NA/0.21 (57, −9, −11)
Insula	13	0.4/0.0	0.21 (−30, 19, 12)/NA
Striatum		0.0/0.2	NA/0.20 (27, 12, 13)
<b>DTI Abbreviation</b>			
DTI Abbreviation	WM tracts	Vo (cc)	Percent
SLF	Superior longitudinal fasciculus	2.1/8.9	2%/9%
UF	Uncinate fasciculus	1/0.2	10%/0
ILF	Inferior longitudinal fasciculus	0.3/0	1%/0
CGH	Cingulum (hippocampus)	0.3/0	3%/0
SLFt	SLF (temporal part)	0/0.3	0/18%
CST	Corticospinal tract	1.8/0	5%/0
FM	Forceps minor/major	3.3/2.1	6%/4%
			Frequency (L/R)
			0.71(15, 27, 37)/0.87(47, 22, 18)
			0.58(16, 45, 9)/NA
			0.58(17, 40, 7)/NA
			0.57(17, 39, 7)/NA
			NA/0.57(45, 21, 15)
			0.48(18, 22, 4)/NA
			0.45(31, 54, 13)/0.48(33, 11, 29)





**Fig. 3.** Identified brain regions that may serve as neuromarkers. Fig. 3 indicates the identified brain clusters of three MRI measures, with the cluster color reflecting its weight and sign, as well as the corresponding regression equation, in which the ultimate correlation with ground truth is 0.7033. In total, 15 clusters were identified, each of which was assigned a weight that implied how much they contribute to the predicted measure (MCCB composite score). The brain views are left, top and front for each column, respectively.

fALFF and  $r = 0.5633$  using only GM). The selected fMRI regions included right MTG (BA 20, 21), parahippocampal gyrus and cerebellum, and sMRI regions included striatum, insula (BA 13); inferior parietal lobule (IPL, BA 39) and angular gyrus (BA 40), which are shown in supplementary Table S4.

#### Prediction of PANSS scores

The PANSS is a medical scale used for measuring symptom severity of patients with schizophrenia (Kay et al., 1987). To evaluate the stability and generality of the proposed framework from another perspective, the PANSS positive and negative scores were also predicted in this project, with  $r = 0.7785$  and  $r = 0.7804$  achieved respectively for all patients, suggesting excellent generalizability of the proposed framework. For both measures, fALFF-GM combination reached the best prediction instead of using three-way fusion (Table 2).

Table 5 provides the identified anatomical labels and occurring frequencies for fALFF and GM. With regard to PANSS positive scores, brain areas including cingulate gyrus, the striatum/thalamus/parahippocampal gyrus and cerebellum were identified in fALFF, whereas the dorsolateral prefrontal cortex (DLPFC, BA10, 46), inferior parietal lobule (BA 39, 40) and cerebellum were identified in GM. While for PANSS negative scores, the striatum, superior frontal gyrus (BA 6) and cerebellum appear significant in both fALFF and GM.

#### Discussion

In this study, we proposed a generalized predicting framework that employs advanced data mining strategies to identify potential neuromarkers by whole brain voxel-wise searching, in which multi-modal MRI information was combined to predict both cognitive and symptomatic scores. In addition, the important brain regions contributing to the regression model were also identified. Both the MCCB composite and PANSS predictions achieved correlations with ground truth  $>0.7$ , and more than 1/3 of the subject samples acquired score estimation with a 90% or higher correlation with true values, which proved the feasibility and the merits of the proposed framework. Note that this prediction framework can easily be generalized to other mental disorders and multiple brain imaging types.

#### Methodological issues

As a data-driven approach, the proposed method has a strength in searching neuromarkers from whole brain voxels within a short time, while most related studies often adopted predefined brain regions as features, such as those based on 116 AAL templates (Wang et al., 2015) or 244 ROIs (Power et al., 2011) to reduce computational load. Though valuable information would be evident using larger ROIs, such methods do not exclude the possibility that some more focused abnormalities could be missed, as reported (Argyelan et al., 2015). By contrast,



**Table 5**

Anatomical information and the frequency of the identified brain ROI features occurring in 47 loops for PANSS prediction.

PANSS Positive $r = 0.7785$			
fMRI_fALFF Area	Brodman area	Vol (cc) L/R	Frequency L/R (x, y, z)
Cingulate gyrus	24	0.2/0.4	0.77 (−12, −6, 27)/0.81 (18, 8, 26)
Striatum		2.5/1.7	0.75 (−12, −12, 19)/0.81 (12, 2, 21)
Medial/Superior frontal Gyrus	10, 11	0.1/0.7	0.11 (−3, 47, −13)/0.70 (18, 50, −6)
Anterior cingulate	34	0.0/0.1	NA (0, 0, 0)/0.70 (15, 47, −6)
Cerebellum		0.0/1.0	NA (0, 0, 0)/0.51 (6, −61, −18)
Parahippocampal gyrus		0.1/0.0	0.51 (−14, 1, −14)/NA
Thalamus		0.6/0.1	0.31 (−9, −27, 9)/0.33 (15, −27, 17)
Precuneus/Cuneus	19, 39	0.6/0.0	0.19 (−24, −78, 42)/NA (0, 0, 0)
sMRI_GM Area	Brodman area	Vol (cc)L/R	Frequency L/R (x, y, z)
Cerebellum		0.5/3.3	0.94/0.95 (6, −64, −24)
Angular gyrus	39	0.2/0.0	0.75 (−36, −61, 35)/NA (0, 0, 0)
Precuneus	19, 39	0.4/0.0	0.75 (−39, −64, 36)/NA (0, 0, 0)
Inferior parietal lobule	39, 40	0.6/0.3	0.75 (−39, −61, 38)/0.75 (54, −35, 40)
Striatum		0.0/0.1	NA (0, 0, 0)/0.72 (12, 16, −4)
Superior/Middle frontal gyrus	46	0.4/0.0	0.65 (−35, 47, −15)/NA
Lingual gyrus	17	0.1/0.3	0.23 (−9, −89, 1)/0.38 (9, −89, 1)
Middle/Inferior frontal gyrus	6, 10	0.6/0.4	0.15 (−36, 37, 19)/0.09 (36, 10, 54)
PANSS Negative $r = 0.7804$			
fMRI_fALFF area	Brodman Area	Vol (cc)L/R	Weights L/R (x, y, z)
Parahippocampal gyrus	34	0.3/0.0	0.98 (−21, 1, −11)/NA (0, 0, 0)
Medial/Superior frontal gyrus	6	0.4/0.0	0.81 (−9, −5, 52)/NA (0, 0, 0)
Cerebellum		0.5/0.0	0.81 (−12, −34, −17)/NA (0, 0, 0)
Cingulate Gyrus	24	0.1/0.0	0.53 (−9, −5, 46)/NA (0, 0, 0)
Striatum		0.9/0.0	0.19 (−12, 13, −1)/NA (0, 0, 0)
sMRI_GM area	Brodman Area	Vol (cc)L/R	Weights L/R (x, y, z)
Striatum		0.0/0.5	NA (0, 0, 0)/0.70 (27, 19, 7)
Superior/Middle frontal gyrus	6, 8, 10	1.8/0.0	0.61 (−29, 24, 57)/NA
Cerebellum		0.4/1.2	0.53 (−24, −64, −20)/0.56 (12, −70, −22)
Angular gyrus/precuneus	39, 40	2.8/0.0	0.56 (−53, −67, 37)/NA
Fusiform/Inferior temporal	20	0.0/0.4	NA (0, 0, 0)/0.33 (57, −29, −19)

we adopted several advanced feature selection approaches, including ReliefF and CFS, to mine from tens of thousands of brain voxels efficiently, and the comparisons in Table 3 did show advantages of the proposed method. Compared with the majority of the heuristic feature selection measures (e.g., information gain (Jakulin et al., 2003), distance measure (Duda et al., 2001)) that assume the conditional independence of attributes, ReliefF is aware of the contextual information; its computational complexity is polynomial and can correctly estimate data sets with dependent and independent attributes (Deng et al., 2010). Regarding parameter tuning, for all feature selection layers, only the ReliefF output needs to be tuned, which can be easily obtained by scree test, see more details in supplementary file. In our case, keeping 2000 ~ 5000 voxels after ReliefF is an optimal and reliable choice. Other feature selection stages can be implemented by WEKA with default settings (<http://www.cs.waikato.ac.nz/ml/weka/index.html>) (Mark Hall et al., 2009). And through the contrast tests in Table 2, we are convinced that each procedure designed for the prediction model adds value in terms of increased performance.

Another point worth noting is that multimodal fusion did improve the prediction performance in both cognitive and symptomatic scores; however, fusing as many modalities as possible in the training sample does not necessarily result in the best predicting results (Sui et al., 2014). To acquire a higher predicting accuracy and make a comparison among combinations of different modalities, we tested all possible modality combinations. As shown in Table 2, three-way fusion is optimal for MCCB composite prediction, whereas for PANSS, ALFF + GM is the best combination since FA map derives very low prediction. Compared to existing neuroimaging prediction studies that focus on single modality or single machine learning algorithms, like the framework proposed by (Mansson et al., 2015), which used SVM to predict long-term treatment response of anxiety disorders via fMRI, or the approach described in (Mwangi et al., 2012), which can predict severity of depression for major depression disorders (MDD) using single MR structural scans, our data-driven framework makes the best use of the multimodal,

sparse, precise yet imperative information, which proves to be flexible and fruitful for an informative understanding of brain activity and disorders (Calhoun and Adali, 2009; Sui et al., 2015; Zhang et al., 2012b).

#### Identified potential neuromarkers

As shown in Table 4, the higher occurring frequency, the greater impact of the brain ROIs contributing to the predicted measure. Note that left middle and superior temporal gyrus (BA 21, 22) have the highest frequency in fMRI and also appears in sMRI. In addition, according to regression equation in Fig. 3, the lower fALFF and higher GM values in BA 21, the higher cognitive performance. This is consistent with findings in (Choi et al., 2012; Kim et al., 2015), where SZs showed dominant activities in the superior and MTG compared with HCs in the implicit memory retrieval tests.

Besides, left middle and medial frontal gyrus occurred in fALFF, which belong to DLPFC and positively correlates with the cognitive scores (Goldberg et al., 2006). Cognitive disorganization has been suggested as one cause of decreased activation in the prefrontal cortex (Perlstein et al., 2001), and DLPFC dysfunction has been related to attention deficits of schizophrenia (Boksman et al., 2005). In addition, abnormal integration of frontal-temporal function, underpinned by a failure of normal cingulate cortical modulation, has also been demonstrated in patients (Fletcher et al., 1999). The significant brain imaging markers located by our regression model are consistent with the above reports.

Furthermore, the cuneus (BA 18, 19), which is most known for its involvement in basic visual processing (Delvecchio et al., 2013; Qiu et al., 2011), also occur in fMRI and sMRI. Since MCCB experiments include visual learning and attention tasks, it makes sense to identify the visual cortex as part of cognitive prediction, and many studies have discovered that SZs present visual processing abnormalities in a variety of tasks (Hardoy et al., 2004; Qiu et al., 2011). Moreover, the similar covariation between fMRI and sMRI suggests a synchronicity and complementary nature between structural and functional changes in cognitive

impairment of schizophrenia, which is supported by (Casey et al., 2005; Salgado-Pineda et al., 2004; Schultz et al., 2012).

With regard to WM assessment, we found the FA mainly lies in parts of WM tracts including the superior longitudinal fasciculus (SLF), UF, ILF, and forceps major/minor, similar to (Wu et al., 2015). Previous results have shown that FA changes in the SLF and SLFt, the major WM connection between the prefrontal and parietal/temporal cortices, relate to verbal working memory performance (Karlsgodt et al., 2008). (Epstein et al., 2014) also found early-onset schizophrenia adolescents exhibit lower FA in left ILF, and the FA values in left ILF predicted worse neurocognitive performance. Similarly, the uncinate fasciculus is suggested as a predictor of conversion from mild cognitive impairment to Alzheimer disease (Hiyoshi-Taniguchi et al., 2015; Serra et al., 2012). All above reports support our findings that the identified WM tracts are associated with or predictive to cognitive impairment.

For MCCB social cognition domain, MTG (BA 20, 21), IPL and precuneus were highly correlated (Supplementary Table S4). IPL, particularly angular gyrus, is concerned with language, mathematical operations, perception of emotions and body image (Radua et al., 2010). Hornak et al. (1996) have reported that damage to the frontal lobes can affect emotional responses to social stimuli and lead to the inability to recognize faces. Furthermore, the striatum and its cortical connections are critical for complex cognition (Radua et al., 2010), and lesions of it can affect various cognitive control processes (Chudasama and Robbins, 2006). All of the identified regions are closely related with the measured MCCB domain social cognition.

When it comes to PANSS, related brain regions with great significance were revealed in striatum, DLPFC, IPL and cerebellum (Table 5). Note that the IPL reveals strong involvement in semantic processing, social cognition or theory-of-mind in several meta-analysis reviews (Binder et al., 2009; Mars et al., 2011; Seghier, 2013). Hazlett, et al. recently reported that larger BA10 volumes in the DLPFC predict less symptom severity (Hazlett et al., 2014), and Kawada et al. found executive dysfunction scores were correlated with volume reduction in DLPFC in schizophrenia (Kawada et al., 2009). The role of the cerebellum in schizophrenia has also been highlighted by Andreasen's hypothesis of "cognitive dysmetria", which suggests a general dyscoordination of sensorimotor and mental processes (Cerasa et al., 2012; Moulis et al., 2014). In addition, thalamus, known for their functions in coordinating, planning (Andreasen et al., 1998) and complex cognition (Simpson et al., 2010), were demonstrated by patients in fMRI. Hence, all of the above brain regions identified in PANSS prediction are closely related to psychotic symptoms in schizophrenia (Makinen et al., 2008; Vohringer et al., 2013).

#### Potential limitations and future directions

There are limitations of this study. First, no clear trend or characteristic was found to suggest which type of regression analysis is optimal, consistent with previous work (Langley and Simon, 1995). Both linear (multivariate linear regression and pace regression) and non-linear (SMOreg) regression were adapted in our study. Further studies are needed to clarify the utility of potential biomarkers in predicting cognitive/symptom performance identified here, which are closely correlated with cognition deficits in schizophrenia. Similar data from two or more sites are ideal for cross-validation. In addition, modalities adopted here can be extended to other types of data, such as genetic data, electroencephalograph (EEG) and PET, etc. Finally, as a general challenge for studying schizophrenia, most participants were receiving antipsychotic and/or mood stabilizing medication during the scanning course, which could result in both structural and functional brain alterations (Ho et al., 2011; Lui et al., 2010). Although we have not examined correlations between antipsychotic dose/gender with the identified brain features, the underlying effects of medication on imaging measures have been widely reported on fALFF (Hadley et al., 2014), DTI (Szeszko

et al., 2014) as well as gray matter (Hutcherson et al., 2014), future studies also need to consider the potential confound of gender effects.

In future work, the proposed method can not only be used for cognitive or symptomatic score prediction, but also be feasible to subcategory classification of mental disorders based on various imaging measures or genetic data. Even more, it can be used for treatment prediction, e.g., predicting that a major depression patient will be a remitter, responder or a non-remitter based on MRI scans before his electroconvulsive therapy, similar to what did in (Johnston et al., 2015; van Waarde et al., 2015), implicating a wide applicability in the neuroimaging field.

In summary, we have developed and evaluated a novel, generalized framework that is able to predict explicit values of targeted measures precisely by using multiple imaging modalities and several superior machine learning approaches. In real application, it was applied to a combination of multiple MRI measures to predict both cognitive and symptomatic scores for schizophrenia and healthy controls. By searching the whole brain voxels, the framework not only achieved relatively higher predicting accuracy for multiple measures, but also demonstrated strong robustness to identify potential biomarkers. In brief, the proposed data-driven framework will be very useful for generating quantitative predictive markers, which can help translate neuroimaging observations into treatment decisions for individualized mental disorders, and thus is possible to lead to correct early intervention and better outcomes.

#### Acknowledgments

This work was partially supported by the National High-Tech Development Plan (863) 2015AA020513, the "100 Talents Plan" of the Chinese Academy of Sciences (to J. Sui); the Chinese National Science Foundation (No.81471367), and the Strategic Priority Research Program of the Chinese Academy of Sciences XDB02060005 (to J. Sui); and the National Institutes of Health grants R01MH084898 (to J. Bustillo); and R01EB006841, and P20GM103472 (to V.D. Calhoun).

The authors report no financial relationship with commercial interests.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.05.026>.

#### References

- Aharoni, E., Vincent, G.M., Harenski, C.L., Calhoun, V.D., Sinnott-Armstrong, W., Gazzaniga, M.S., Kiehl, K.A., 2013. Neuroprediction of future rearrest. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6223–6228.
- Allen, E.A., Erhardt, E.B., Damaraju, E., Gruner, W., Segall, J.M., Silva, R.F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., Michael, A.M., Caprihan, A., Turner, J.A., Eichele, T., Adelsheim, S., Bryan, A.D., Bustillo, J., Clark, V.P., Feldstein Ewing, S.W., Filbey, F., Ford, C.C., Hutchison, K., Jung, R.E., Kiehl, K.A., Koditwakkul, P., Komesu, Y.M., Mayer, A.R., Pearson, G.D., Phillips, J.P., Sadek, J.R., Stevens, M., Teuscher, U., Thoma, R.J., Calhoun, V.D., 2011. A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5, 2.
- Andreasen, N.C., Paradiso, S., O'Leary, D.S., 1998. "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? *Schizophr. Bull.* 24, 203–218.
- Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *NeuroImage* 145, 137–165.
- Argyelan, M., Gallego, J.A., Robinson, D.G., Ikuta, T., Sarpal, D., John, M., Kingsley, P.B., Kane, J., Malhotra, A.K., Szeszko, P.R., 2015. Abnormal resting state fMRI activity predicts processing speed deficits in first-episode psychosis. *Neuropsychopharmacology* 40, 1631–1639.
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *NeuroImage* 26, 839–851.
- August, S.M., Kiwanuka, J.N., McMahon, R.P., Gold, J.M., 2011. The MATRICS Consensus Cognitive Battery (MCCB): clinical and cognitive correlates. *Schizophr. Res.* 134, 76–82.
- Bazin, P.-L., Shiee, N., Ellingsen, L.M., Prince, J.L., Pham, D.L., 2011. Digital Topology in Brain Image Segmentation and Registration. In: El-Baz, A.S., Acharya, U.R., Mirmehdi, M., Suri, J.S. (Eds.), *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*. Springer, pp. 339–375.
- Bielza, C., Larranaga, P., 2014. Discrete Bayesian network classifiers: a survey. *ACM Comput. Surv.* 47.

- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.
- Boksman, K., Theberge, J., Williamson, P., Drost, D.J., Malla, A., Densmore, M., Takhar, J., Pavlosky, W., Menon, R.S., Neufeld, R.W., 2005. A 4.0-T fMRI study of brain connectivity during word fluency in first-episode schizophrenia. *Schizophr. Res.* 75, 247–263.
- Burton, C.Z., Vella, L., Harvey, P.D., Patterson, T.L., Heaton, R.K., Twamley, E.W., 2013. Factor structure of the MATRICS Consensus Cognitive Battery (MCCB) in schizophrenia. *Schizophr. Res.* 146, 244–248.
- Calhoun, V.D., Adali, T., 2009. Feature-based fusion of medical imaging data. *IEEE Trans. Inf. Technol. Biomed.* 13, 711–720.
- Calhoun, V.D., Allen, E., 2012. Extracting Intrinsic Functional Networks with Feature-based Group Independent Component Analysis. *Psychometrika* 1–17 (in press).
- Calhoun, V.D., Sui, J., 2016. Multimodal fusion of brain imaging data: a key to finding the missing link(s) in complex mental illness. *Biol. Psychiatry: Cogn. Neurosci. Neuroimaging* (in press).
- Casey, B.J., Tottenham, N., Liston, C., Durston, S., 2005. Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* 9, 104–110.
- Cerasa, A., Cherubini, A., Peran, P., 2012. Multimodal MRI in neurodegenerative disorders. *Neurol. Res. Int.* 2012, 287891.
- Chen, C.H., Ridler, K., Suckling, J., Williams, S., Fu, C.H., Merlo-Pich, E., Bullmore, E., 2007. Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biol. Psychiatry* 62, 407–414.
- Cheng, C.C., Peng, G.J., Hwang, W.L., 2009. Subband weighting with pixel connectivity for 3-D wavelet coding. *IEEE Trans. Image Process.* 18, 52–62.
- Choi, J.S., Park, J.Y., Jung, M.H., Jang, J.H., Kang, D.H., Jung, W.H., Han, J.Y., Choi, C.H., Hong, K.S., Kwon, J.S., 2012. Phase-specific brain change of spatial working memory processing in genetic and ultra-high risk groups of schizophrenia. *Schizophr. Bull.* 38, 1189–1199.
- Chudasama, Y., Robbins, T.W., 2006. Functions of frontostriatal systems in cognition: comparative neuropsychopharmacological studies in rats, monkeys and humans. *Biol. Psychol.* 73, 19–38.
- Clark, V.P., Beatty, G.K., Anderson, R.E., Kodituwakku, P., Phillips, J.P., Lane, T.D., Kiehl, K.A., Calhoun, V.D., 2014. Reduced fMRI activity predicts relapse in patients recovering from stimulant dependence. *Hum. Brain Mapp.* 35, 414–428.
- Cuthbert, B.N., Workgroup, N.R., 2014. The RDoC framework: continuing commentary. *World Psychiatry* 13, 196–197.
- Delvecchio, G., Sugranyes, G., Frangou, S., 2013. Evidence of diagnostic specificity in the neural correlates of facial affect processing in bipolar disorder and schizophrenia: a meta-analysis of functional imaging studies. *Psychol. Med.* 43, 553–569.
- Deng, Z.H., Chung, F.L., Wang, S.T., 2010. Robust relief-feature weighting, margin maximization, and fuzzy optimization. *IEEE Trans. Fuzzy Syst.* 18, 726–744.
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. second ed. Wiley, New York; Chichester.
- Dumpuri, P., Thompson, R.C., Cao, A., Ding, S., Garg, I., Dawant, B.M., Miga, M.I., 2010. A fast and efficient method to compensate for brain shift for tumor resection therapies measured between preoperative and postoperative tomograms. *IEEE Trans. Biomed. Eng.* 57, 1285–1296.
- Edwards, L., Perrin, K., Wijesinghe, M., Weatherall, M., Beasley, R., Travers, J., 2011. The value of the CRB65 score to predict mortality in exacerbations of COPD requiring hospital admission. *Respirology* 16, 625–629.
- Eichele, T., Debener, S., Calhoun, V.D., Specht, K., Engel, A.K., Hugdahl, K., von Cramon, D.Y., Ullsperger, M., 2008. Prediction of human errors by maladaptive changes in event-related brain networks. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6173–6178.
- Epstein, K.A., Cullen, K.R., Mueller, B.A., Robinson, P., Lee, S., Kumra, S., 2014. White matter abnormalities and cognitive impairment in early-onset schizophrenia-spectrum disorders. *J. Am. Acad. Child Adolesc. Psychiatry* 53 (362–372), e361–e362.
- Fletcher, P., McKenna, P.J., Friston, K.J., Frith, C.D., Dolan, R.J., 1999. Abnormal cingulate modulation of fronto-temporal connectivity in schizophrenia. *NeuroImage* 9, 337–342.
- Frank, E., Hall, M., Trigg, L., Holmes, G., Witten, I.H., 2004. Data mining in bioinformatics using Weka. *Bioinformatics* 20, 2479–2481.
- Freire, L., Roche, A., Mangin, J.F., 2002. What is the best similarity measure for motion correction in fMRI time series? *IEEE Trans. Med. Imaging* 21, 470–484.
- Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.P., Heather, J.D., Frackowiak, R.S., 1995. Spatial registration and normalization of images. *Hum. Brain Mapp.* 2, 165–189.
- Gabrieli, J.D., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85, 11–26.
- Goldberg, I.L., Harel, M., Malach, R., 2006. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339.
- Gudmundsson, S., Runarsson, T.P., Sigurdsson, S., 2012. Test-retest reliability and feature selection in physiological time series classification. *Comput. Methods Prog. Biomed.* 105, 50–60.
- Hadley, J.A., Nenert, R., Kraguljac, N.V., Bolding, M.S., White, D.M., Skidmore, F.M., Visscher, K.M., Lahti, A.C., 2014. Ventral tegmental area/midbrain functional connectivity and response to antipsychotic medication in schizophrenia. *Neuropsychopharmacology* 39, 1020–1030.
- Hardoy, M.C., Carta, M.G., Catena, M., Hardoy, M.J., Cadeddu, M., Dell'Osso, L., Hugdahl, K., Carpiniello, B., 2004. Impairment in visual and spatial perception in schizophrenia and delusional disorder. *Psychiatry Res.* 127, 163–166.
- Hazlett, E.A., Lamade, R.V., Graff, F.S., McClure, M.M., Kolaitis, J.C., Goldstein, K.E., Siever, L.J., Godbold, J.H., Moshier, E., 2014. Visual-spatial working memory performance and temporal gray matter volume predict schizotypal personality disorder group membership. *Schizophr. Res.* 152, 350–357.
- Hiyoshi-Taniguchi, K., Oishi, N., Namiki, C., Miyata, J., Murai, T., Cichocki, A., Fukuyama, H., 2015. The uncinate fasciculus as a predictor of conversion from amnesic mild cognitive impairment to Alzheimer disease. *J. Neuroimaging* 25, 748–753.
- Ho, B.C., Andreasen, N.C., Ziebell, S., Pierson, R., Magnotta, V., 2011. Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. *Arch. Gen. Psychiatry* 68, 128–137.
- Hornak, J., Rolls, E.T., Wade, D., 1996. Face and voice expression identification in patients with emotional and behavioral changes following ventral frontal lobe damage. *Neuropsychologia* 34 (4), 247–261.
- Hutcherson, N.L., Clark, D.G., Bolding, M.S., White, D.M., Lahti, A.C., 2014. Basal ganglia volume in unmedicated patients with schizophrenia is associated with treatment response to antipsychotic medication. *Psychiatry Res.* 221, 6–12.
- Hutton, C., Draganski, B., Ashburner, J., Weiskopf, N., 2009. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *NeuroImage* 48, 371–380.
- Jakulin, A., Bratko, I., Smrke, D., Demsar, J., Zupan, B., 2003. Attribute interactions in medical data analysis. *Artif. Intell. Med. Proc.* 2780, 229–238.
- Johnston, B.A., Steele, J.D., Tolomeo, S., Christmas, D., Matthews, K., 2015. Structural MRI-Based Predictions in Patients with Treatment-Refractory Depression (TRD). *PLoS One* 10, e0132958.
- Karlsgodt, K.H., van Erp, T.G., Poldrack, R.A., Bearden, C.E., Nuechterlein, K.H., Cannon, T.D., 2008. Diffusion tensor imaging of the superior longitudinal fasciculus and working memory in recent-onset schizophrenia. *Biol. Psychiatry* 63, 512–518.
- Kawada, R., Yoshizumi, M., Hirao, K., Fujiwara, H., Miyata, J., Shimizu, M., Namiki, C., Sawamoto, N., Fukuyama, H., Hayashi, T., Murai, T., 2009. Brain volume and dysexecutive behavior in schizophrenia. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 33, 1255–1260.
- Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276.
- Kim, G.W., Yang, J.C., Jeong, G.W., 2015. Emotional effect on cognitive control in implicit memory tasks in patients with schizophrenia. *Neuroreport* 26, 647–655.
- Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *Artif. Intell.* 97, 273–324.
- Langley, P., Simon, H.A., 1995. Applications of machine learning and rule induction. *Commun. ACM* 38, 55–64.
- Leergaard, T.B., Hilgetag, C.C., Sporns, O., 2012. Mapping the connectome: multi-level analysis of brain connectivity. *Front. Neuroinform.* 6, 14.
- Li, C., Jiang, L., 2006. Using Locally Weighted Learning to Improve Smoreg for Regression. *Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence*. Springer-Verlag, Guilin, China, pp. 375–384.
- Lui, S., Li, T., Deng, W., Jiang, L., Wu, Q., Tang, H., Yue, Q., Huang, X., Chan, R.C., Collier, D.A., Meda, S.A., Pearson, G., Mechelli, A., Sweeney, J.A., Gong, Q., 2010. Short-term effects of antipsychotic treatment on cerebral function in drug-naïve first-episode schizophrenia revealed by “resting state” functional magnetic resonance imaging. *Arch. Gen. Psychiatry* 67, 783–792.
- Makinen, J., Miettinen, J., Isohanni, M., Koponen, H., 2008. Negative symptoms in schizophrenia: a review. *Nord. J. Psychiatry* 62, 334–341.
- Mansson, K.N., Frick, A., Boraxbekk, C.J., Marquand, A.F., Williams, S.C., Carlbring, P., Andersson, G., Furmark, T., 2015. Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl. Psychiatry* 5, e530.
- Mark Hall, E.F., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. *SIGKDD Explor.* 11.
- Mars, R.B., Jbabdi, S., Sallet, J., O'Reilly, J.X., Croxson, P.L., Olivier, E., Noonan, M.P., Bergmann, C., Mitchell, A.S., Baxter, M.G., Behrens, T.E., Johansen-Berg, H., Tomassini, V., Miller, K.L., Rushworth, M.F., 2011. Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J. Neurosci.* 31, 4087–4100.
- Meshkin, A., Sadeghi, M., Ghasem-Aghaee, N., 2009. Prediction of Relative Solvent Accessibility Using Pace Regression. *Excli J.* 8, 211–217.
- Mohammadi, P., Hatamlou, A., Masdari, M., 2013. A comparative study on remote tracking of Parkinson's disease progression using data mining methods. *CORR* 3.
- Mori, K., Hasegawa, J., Suenaga, Y., Toriwaki, J., 2000. Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Trans. Med. Imaging* 19, 103–114.
- Moulis, G., Sailer, L., Bonneville, F., Wagner, T., 2014. Imaging in Erdheim-Chester disease: classic features and new insights. *Clin. Exp. Rheumatol.* 32, 410–414.
- Mwangi, B., Matthews, K., Steele, J.D., 2012. Prediction of illness severity in patients with major depression using structural MR brain scans. *J. Magn. Reson. Imaging* 35, 64–71.
- Natsopoulos, D., Kiosseoglou, G., Xeromeritou, A., Alevriadou, A., 1998. Do the hands talk on mind's behalf? Differences in language ability between left- and right-handed children. *Brain Lang.* 64, 182–214.
- Oakes, T.R., Fox, A.S., Johnstone, T., Chung, M.K., Kalin, N., Davidson, R.J., 2007. Integrating VBM into the General Linear Model with voxelwise anatomical covariates. *NeuroImage* 34, 500–508.
- Oto, A., Yang, C., Kayhan, A., Tretiakova, M., Antic, T., Schmid-Tannwald, C., Eggner, S., Karcmar, G.S., Stadler, W.M., 2011. Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. *AJ. Am. J. Roentgenol.* 197, 1382–1390.
- Perlstein, W.M., Carter, C.S., Noll, D.C., Cohen, J.D., 2001. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am. J. Psychiatry* 158, 1105–1113.
- Platt, J.C., 1999. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: Bernhard, S., Ikonf Christopher, J.C.B., Alexander, J.S. (Eds.), *Advances in Kernel Methods*. MIT Press, pp. 185–208.
- Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional network organization of the human brain. *Neuron* 72, 665–678.



- Qin, J., Chen, S.-G., Hu, D., Zeng, L.-L., Fan, Y.-M., Chen, X.-P., Shen, H., 2015. Predicting individual brain maturity using dynamic functional connectivity. *Front. Hum. Neurosci.* 9.
- Qiu, L., Tian, L., Pan, C., Zhu, R., Liu, Q., Yan, J., Zhao, Q., Yuan, H., Han, Y., Yue, W., Yan, H., Zhang, D., 2011. Neuroanatomical circuitry associated with exploratory eye movement in schizophrenia: a voxel-based morphometric study. *PLoS One* 6, e25805.
- Radua, J., Phillips, M.L., Russell, T., Lawrence, N., Marshall, N., Kalidindi, S., El-Hage, W., McDonald, C., Giampietro, V., Brammer, M.J., David, A.S., Surguladze, S.A., 2010. Neural response to specific components of fearful faces in healthy and schizophrenic adults. *NeuroImage* 49, 939–946.
- Radua, J., Canales-Rodriguez, E.J., Pomarol-Clotet, E., Salvador, R., 2014. Validity of modulation and optimal settings for advanced voxel-based morphometry. *NeuroImage* 86, 81–90.
- Robnik-Šikonja, M., Kononenko, I., 1997. An Adaptation of Relief for Attribute Estimation in Regression. In: Fisher, D.H. (Ed.), *ICML*. Morgan Kaufmann, pp. 296–304.
- Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* 53, 23–69.
- Salgado-Pineda, P., Junque, C., Vendrell, P., Baeza, I., Bargallo, N., Falcon, C., Bernardo, M., 2004. Decreased cerebral activation during CPT performance: structural and functional deficits in schizophrenic patients. *NeuroImage* 21, 840–847.
- Schölkopf, B., Smola, A.J., 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, Mass.
- Schultz, C.C., Fusar-Poli, P., Wagner, G., Koch, K., Schachtzabel, C., Gruber, O., Sauer, H., Schlosser, R.G., 2012. Multimodal functional and structural imaging investigations in psychosis research. *Eur. Arch. Psychiatry Clin. Neurosci.* 262 (Suppl. 2), S97–106.
- Segall, J.M., Turner, J.A., van Erp, T.G., White, T., Bockholt, H.J., Gollub, R.L., Ho, B.C., Magnotta, V., Jung, R.E., McCarley, R.W., Schulz, S.C., Lauriello, J., Clark, V.P., Voyvodic, J.T., Diaz, M.T., Calhoun, V.D., 2009. Voxel-based morphometric multisite collaborative study on schizophrenia. *Schizophr. Bull.* 35, 82–95.
- Seghier, M.L., 2013. The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19, 43–61.
- Serra, L., Cercignani, M., Basile, B., Spano, B., Perri, R., Fadda, L., Marra, C., Giubilei, F., Caltagirone, C., Bozzali, M., 2012. White matter damage along the uncinate fasciculus contributes to cognitive decline in AD and DLB. *Curr. Alzheimer Res.* 9, 326–333.
- Shaffer, J.L., Petrella, J.R., Sheldon, F.C., Choudhury, K.R., Calhoun, V.D., Coleman, R.E., Doraiswamy, P.M., Alzheimer's Disease Neuroimaging, I., 2013. Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* 266, 583–591.
- Shen, L., Thompson, P.M., Potkin, S.G., Bertram, L., Farrer, L.A., Foroud, T.M., Green, R.C., Hu, X., Huentelman, M.J., Kim, S., Kauwe, J.S., Li, Q., Liu, E., Macciardi, F., Moore, J.H., Munsie, L., Nho, K., Ramanan, V.K., Risacher, S.L., Stone, D.J., Swaminathan, S., Toga, A.W., Weiner, M.W., Saykin, A.J., 2014. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.* 8, 183–207.
- Simpson, E.H., Kellendonk, C., Kandel, E., 2010. A possible role for the striatum in the pathogenesis of the cognitive symptoms of schizophrenia. *Neuron* 65, 585–596.
- Simpson, A.L., Geller, D.A., Hemming, A.W., Jarnagin, W.R., Clements, L.W., D'Angelica, M.I., Dumpuri, P., Gonen, M., Zendejas, I., Miga, M.I., Stefansic, J.D., 2014. Liver planning software accurately predicts postoperative liver volume and measures early regeneration. *J. Am. Coll. Surg.* 219, 199–207.
- Stokes, M.E., Visweswaran, S., 2012. Application of a spatially-weighted Relief algorithm for ranking genetic predictors of disease. *BioData Min.* 5, 20.
- Stonington, C.M., Chu, C., Kloppel, S., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S., Alzheimer Disease Neuroimaging, I., 2010. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *NeuroImage* 51, 1405–1413.
- Sui, J., Pearson, G., Caprihan, A., Adali, T., Kiehl, K.A., Liu, J., Yamamoto, J., Calhoun, V.D., 2011. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA + joint ICA model. *NeuroImage* 57, 839–855.
- Sui, J., Adali, T., Yu, Q., Chen, J., Calhoun, V.D., 2012. A review of multivariate methods for multimodal fusion of brain imaging data. *J. Neurosci. Methods* 204, 68–81.
- Sui, J., He, H., Pearson, G.D., Adali, T., Kiehl, K.A., Yu, Q., Clark, V.P., Castro, E., White, T., Mueller, B.A., Ho, B.C., Andreasen, N.C., Calhoun, V.D., 2013. Three-way (N-way) fusion of brain imaging data based on mCCA + jICA and its application to discriminating schizophrenia. *NeuroImage* 66, 119–132.
- Sui, J., Huster, R., Yu, Q., Segall, J.M., Calhoun, V.D., 2014. Function–structure associations of the brain: evidence from multimodal connectivity and covariance studies. *NeuroImage* 102P1, 11–23.
- Sui, J., Pearson, G.D., Du, Y., Yu, Q., Jones, T.R., Chen, J., Jiang, T., Bustillo, J., Calhoun, V.D., 2015. In Search of multimodal neuroimaging biomarkers of cognitive deficits in schizophrenia. *Biol. Psychiatry*.
- Sun, Y.J., 2007. Iterative RELIEF for feature weighting: algorithms, theories, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 1035–1051.
- Szeszko, P.R., Robinson, D.G., Ikuta, T., Peters, B.D., Gallego, J.A., Kane, J., Malhotra, A.K., 2014. White matter changes associated with antipsychotic treatment in first-episode psychosis. *Neuropsychopharmacology* 39, 1324–1331.
- Technow, J.R., Hazel, N.A., Abela, J.R., Hankin, B.L., 2014. Stress sensitivity interacts with depression history to predict depressive symptoms among youth: prospective changes following first depression onset. *J. Abnorm. Child Psychol.*
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 267–288.
- Tibshirani, R.J., 2011. *The Solution Path of the Generalized Lasso*. Stanford University.
- Tripoliti, E.E., Fotiadis, D.I., Argyropoulou, M., Manis, G., 2010. A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. *J. Biomed. Inform.* 43, 307–320.
- Tsang, J.M., Dougherty, R.F., Deutsch, G.K., Wandell, B.A., Ben-Shachar, M., 2009. Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22546–22551.
- van Waarde, J.A., Scholte, H.S., van Oudheusden, L.J., Verwey, B., Denys, D., van Wingen, G.A., 2015. A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Mol. Psychiatry* 20, 609–614.
- Vittengl, J.R., Clark, L.A., Thase, M.E., Jarrett, R.B., 2014. Replication and extension: separate personality traits from states to predict depression. *J. Personal. Disord.* 28, 225–246.
- Vohringer, P.A., Barroilhet, S.A., Amerio, A., Reale, M.L., Alvear, K., Vergne, D., Ghaemi, S.N., 2013. Cognitive impairment in bipolar disorder and schizophrenia: a systematic review. *Front. Psychol.* 4, 87.
- Wan, J., Zhang, Z.L., Rao, B.D., Fang, S.F., Yan, J.W., Saykin, A.J., Shen, L., Neuroimaging, A.S.D., 2014. Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation- and nonlinearity-aware sparse Bayesian learning. *IEEE Trans. Med. Imaging* 33, 1475–1487.
- Wang, Y., 2000. *A New Approach to Fitting Linear Models in High Dimensional Spaces*. University of Waikato (pp. xiii, 204 leaves).
- Wang, H., Nie, F., Huang, H., Risacher, S., Ding, C., Saykin, A.J., Shen, L., Adni, 2011a. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. *Proc. IEEE Int. Conf. Comput. Assist. Interv.* 14, 115–123.
- Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, A.J., Shen, L., 2011b. Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. *Med. Image Comput. Comput. Assist. Interv.* 14, 115–123.
- Wang, J., Cao, H., Liao, Y., Liu, W., Tan, L., Tang, Y., Chen, J., Xu, X., Li, H., Luo, C., Liu, C., Ries Merikangas, K., Calhoun, V., Tang, J., Shugart, Y.Y., Chen, X., 2015. Three dysconnectivity patterns in treatment-resistant schizophrenia patients and their unaffected siblings. *NeuroImage Clin.* 8, 95–103.
- White, T., O'Leary, D., Magnotta, V., Arndt, S., Flaum, M., Andreasen, N.C., 2001. Anatomic and functional variability: the effects of filter size in group fMRI data analysis. *NeuroImage* 13, 577–588.
- Wilhelm-Benartzi, C.S., Koestler, D.C., Karagas, M.R., Flanagan, J.M., Christensen, B.C., Kelsey, K.T., Marsit, C.J., Houseman, E.A., Brown, R., 2013. Review of processing and analysis methods for DNA methylation array data. *Br. J. Cancer* 109, 1394–1402.
- Willette, A.A., Calhoun, V.D., Egan, J.M., Kapogiannis, D., Alzheimer's Disease Neuroimaging, I., 2014. Prognostic classification of mild cognitive impairment and Alzheimer's disease: MRI independent component analysis. *Psychiatry Res.* 224, 81–88.
- Wu, L., Calhoun, V.D., Jung, R.E., Caprihan, A., 2015. Connectivity-based whole brain dual parcellation by group ICA reveals tract structures and decreased connectivity in schizophrenia. *Hum. Brain Mapp.* 36, 4681–4701.
- Yan, J., Huang, H., Risacher, S.L., Kim, S., Inlow, M., Moore, J.H., Saykin, A.J., Shen, L., 2013. Network-guided sparse learning for predicting cognitive outcomes from MRI measures. *Multimodal Brain Image Anal.* 8159, 202–210 (2013).
- Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., Novak, G., DiBernardo, A., Narayan, V.A., Alzheimer's Disease Neuroimaging, I., 2012. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurol.* 12, 46.
- Zhang, D., Shen, D., Alzheimer's Disease Neuroimaging, I., 2012a. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage* 59, 895–907.
- Zhang, H., Liu, L., Wu, H., Fan, Y., 2012b. Feature Selection and SVM Classification of Multiple Modality Images for Predicting MCI. *OHBM Annual Meeting*.