



Combining Multiple Resting-State fMRI Features during Classification: Optimized Frameworks and Their Application to Nicotine Addiction

Xiaoyu Ding, Yihong Yang, Elliot A. Stein and Thomas J. Ross*

Neuroimaging Research Branch, Intramural Research Program, National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Satrajit S. Ghosh,
Massachusetts Institute of
Technology, United States

Reviewed by:

Xin Di,
New Jersey Institute of Technology,
United States
Arun Bokde,
Trinity College, Dublin, Ireland

*Correspondence:

Thomas J. Ross
tross@mail.nih.gov

Received: 16 March 2017

Accepted: 26 June 2017

Published: 12 July 2017

Citation:

Ding X, Yang Y, Stein EA and Ross TJ
(2017) Combining Multiple
Resting-State fMRI Features during
Classification: Optimized Frameworks
and Their Application to Nicotine
Addiction.
Front. Hum. Neurosci. 11:362.
doi: 10.3389/fnhum.2017.00362

Machine learning techniques have been applied to resting-state fMRI data to predict neurological or neuropsychiatric disease states. Existing studies have used either a single type of resting-state feature or a few feature types (<4) in the prediction model. However, resting-state data can be processed in many different ways, yielding different feature types containing complementary and/or novel information, leaving uncertain the most informative features to provide to the classifier. In this study, multiple resting-state features were calculated from two main analytical categories: local measures and network measures. Feature selection was adopted using an optimized grid-search approach selecting top ranked features from statistical tests. We then tested three optimized frameworks: feature combination, kernel combination, and classifier combination, all using the support vector machine as an elementary classifier, to combine these resting-state feature types. When applied to nicotine addiction, with a cohort size of 100 smokers and 100 non-smokers, via a 10-fold cross-validation procedure, the feature combination and the classifier combination achieved an accuracy of 75.5%, while the kernel combination achieved a 73.0% accuracy; all three combination frameworks improved classification performance compared to the single feature type based results (best accuracy 70.5%). This study not only reveals the discriminative power of resting-state data, but also demonstrates the efficiency of combining multiple features from one data phenotype to improve classification performance.

Keywords: feature combination, kernel combination, classifier combination, resting-state fMRI, nicotine addiction, support vector machine

INTRODUCTION

Machine learning techniques are playing an increasingly important role in neuroscience research to explore various brain functions (Klöppel et al., 2012; Richiardi et al., 2013; Sundermann et al., 2014; Gabrieli et al., 2015). They have been applied to neuroimaging data to predict group membership, which may lead to brain-based biomarkers of disease (Chen and Herskovits, 2010; Wang et al., 2010; Zhang and Shen, 2012; Hart et al., 2014; Pariyadath et al., 2014; Ding et al., 2015; Jie et al., 2015; Libero et al., 2015; Liu et al., 2015; Moradi et al., 2015; Suk et al., 2015; Arbabshirani et al., 2016). A prominent advantage of machine learning algorithms is that they learn a computational model from exemplar inputs, which can later be applied to new unknown samples to make predictions or

decisions. Moreover, discriminative features selected by machine learning techniques can uncover multivariate relationships beyond those found by univariate analysis such as simple statistical tests. For neuroimaging data, the model is usually evaluated using a cross-validation (CV) procedure, in which a set of data is split into complementary subsets separately used for training and testing the model (Hirsch, 1991; Wolfers et al., 2015).

Support vector machine (SVM) is one of the most popular machine learning algorithms that has been applied to neuroimaging data (for a review, see Orrù et al., 2012). In binary classification, given a set of training samples, each with a label marked for its category, a SVM constructs a separating hyperplane that maximizes the margin between samples (Cortes and Vapnik, 1995; Burges, 1998). However, frequently the sets are not linearly separable in the original input space. In this case, these samples are first mapped into a higher dimensional space using a kernel function, which presumably makes the separation easier in the transformed space. A commonly adopted kernel function is the Gaussian radial basis function (RBF) that maps the input samples into a Hilbert space, corresponding to a non-linear SVM called RBF kernel SVM (RBF-SVM) (Burges, 1998).

Resting-state fMRI is a functional brain imaging method that measures spontaneous fluctuations in blood-oxygen-level dependent (BOLD) signals that occur in the absence of an explicit task (for a review, see Lee et al., 2013). Thus, the resting-state approach is ideal to examine brain function in patients who may experience difficulty in performing tasks. Resting-state data are widely investigated using machine learning approaches (Deshpande et al., 2010; Shen et al., 2010; Dai et al., 2012; Eloyan et al., 2012; Zeng et al., 2014; Iidaka, 2015; Liu et al., 2015; Rehme et al., 2015). For example, our group previously applied SVM-based classification to resting-state functional connectivity (rsFC) data from 21 smokers and 21 non-smokers to successfully predict smoking status (Pariyadath et al., 2014). Three network characteristics, including network representativeness, within network connectivity, and between network connectivity were tested, separately. Among these, within network connectivity offered maximal information for predicting smoking status with an accuracy of 78.6% using leave-one-out cross-validation (LOOCV).

As in the above example, most studies that applied machine learning techniques to resting-state data used either a single type of resting-state feature (Shen et al., 2010; Zeng et al., 2014; Iidaka, 2015; Liu et al., 2015; Rehme et al., 2015) or just a few feature types (<4) (Deshpande et al., 2010; Dai et al., 2012) to do prediction. However, resting-state data can be processed in many different ways, yielding different feature types containing complementary and/or novel information. When applied to brain disorders using machine learning, these feature types may also provide disparate discriminative information. Furthermore, they may be combined in different ways, potentially leading to an improved model performance.

The purpose of the present methodological study was to determine the optimal resting-state feature types to enter into several classification models. Multiple resting-state feature

types were calculated from two main data categories: local measures and network measures. Feature selection was adopted using an optimized grid-search approach selecting top ranked features from two-sample *t*-tests. We then implemented three optimized classification frameworks: feature combination, kernel combination, and classifier combination, all using the RBF-SVM as an elementary classifier.

MATERIALS AND METHODS

Participants

In order to evaluate their performance, the three frameworks were applied to existing nicotine addiction data from our lab. One hundred cigarette smokers and 100 non-smoking healthy control participants matched on age and gender (see **Table 1** for demographics) were enrolled under several protocols approved by the Institutional Review Board of the National Institute on Drug Abuse Intramural Research Program (NIDA-IRP). Smokers were not currently trying to quit or seeking smoking cessation treatment and were allowed to smoke *ad libitum* prior to the scan session. Controls were included if they had smoked fewer than 25 cigarettes in their lifetime and none in the past year. Potential participants were assessed with a comprehensive medical history and physical exam, general urine and blood laboratory panels, a computerized Structured Clinical Interview for DSM-IV with follow-up clinical interview, and a drug use survey. Participants were excluded if they had any major medical illness, history of neurological or psychiatric disorders, or current or past dependence on any drug other than nicotine. All participants provided written informed consent approved by the NIDA-IRP IRB and received monetary compensation for their participation.

Data Acquisition and Preprocessing

Functional MRI data were collected at the NIDA-IRP on a 3T Siemens Allegra MRI scanner (Erlangen, Germany) equipped with a standard radio frequency birdcage head coil. During the resting-state scanning, 39 slices, without interslice gap, 30° from AC-PC, were prescribed to cover the whole brain. The resting-state data were acquired using a single-shot gradient

TABLE 1 | Demographics of the participants.

	Smokers	Non-smokers
Number	100	100
Age	31.9 ± 9.5	32.6 ± 9.9
Gender	53 M, 47 F	53 M, 47 F
FTND	5.3 ± 1.9	–
CPD	18.2 ± 6.8	–
Smoking years	15.3 ± 9.4	–
Lifetime usage	14.5 ± 11.9	–

P > 0.6 on age between smokers and non-smokers.

FTND: Fagerström Test for Nicotine Dependence.

CPD: cigarettes per day.

Lifetime usage: measured in pack-years (= CPD × smoking years/20).

M/F: male/female.

Age, FTND, CPD, smoking years and lifetime usage are calculated in mean ± SD.

echo-planar imaging (EPI) sequence with repetition time (TR) of 2,000 ms, echo time (TE) of 27 ms, flip angle (FA) = 80°, field of view (FOV) of 220 × 220 mm, and acquisition matrix of 64 × 64, resulting in 300 volumes for each subject. For registration purposes, high-resolution anatomical images were acquired using a 3D magnetization prepared rapid gradient-echo (MPRAGE) T1-weighted sequence in 1 mm³ isotropic voxels (TR = 2,500 ms, TE = 4.38 ms, FA = 8°).

Data preprocessing were conducted in AFNI (Cox, 1996) including slice timing and head motion correction. Data were then spatially normalized to a template in Talairach space to a resampled resolution of 3 × 3 × 3 mm³. White matter (WM) and cerebrospinal fluid (CSF) signals, originating presumably from such systemic effects as respiration and cardiac-induced pulsations, were accounted for individually by extracting the first three principal components from a WM time course ensemble and the first three principal components from a CSF time course ensemble (Behzadi et al., 2007). Here the WM and CSF masks were generated by segmenting the high resolution structural images in AFNI (3dSeg) and down sampling the obtained WM and CSF masks to the same resolution as the functional data. In addition to these physiological regressors, time courses of the six motion parameters also served as uninteresting covariates. The data were temporally band-pass filtered (0.01–0.1 Hz) and uninteresting covariates were removed simultaneously using 3dBandpass in AFNI. Next the data were spatially smoothed with an 8 mm full-width half-maximum (FWHM) Gaussian kernel to increase spatial signal to noise ratio. Finally, data were censored for motion with a threshold of 0.35 for a frame-to-frame change in Euclidean norm of the six motion parameters (Power et al., 2012, 2014, 2015). Two smokers whose censored volumes exceeded 1/3 of the original time series were removed from further analysis (The excluded subjects are not included in **Table 1**). All remaining subjects had at least 80% of their data retained. Further, there was no significant difference ($p = 0.995$) between groups on the numbers of time points that were censored. Most of the feature sets were calculated using the censored data, however, as noted in the following section, some of them utilized the uncensored data.

Feature Extraction

Since resting-state data can be processed in many different ways yielding different feature types, we calculated multiple resting-state feature types from two main analysis categories: local measures and network measures. These feature types are detailed below. Our motivation for extracting these feature types were 2-fold: first, these are among the most common ways that resting-state data are analyzed; second, it has been demonstrated that nicotine dependent individuals show abnormalities in many of these resting-state features, with the expectation therefore of maximizing our ability to separate the groups (Sutherland et al., 2012; Ding and Lee, 2013; Fedota and Stein, 2015; Wu et al., 2015).

Local Measures

Four local measures including the amplitude of low frequency fluctuations, regional homogeneity, voxel-mirrored homotopic

connectivity, and functional connectivity strength, were calculated. We consider these as local measures given that they represent values in each brain region (i.e., nodes in graph theory).

Amplitude of Low Frequency Fluctuations (ALFF)

ALFF measures regional spontaneous fluctuations in BOLD signal intensity in the resting-state brain. Briefly, the time series of preprocessed but uncensored data was transformed to the frequency domain using a fast Fourier transform (FFT). The square root of the power spectrum was calculated at each frequency and then averaged across 0.01–0.1 Hz at each voxel. This averaged square root was taken as the ALFF (Zang et al., 2007). For standardization purpose (i.e., reducing the global effects of variability across subjects), the ALFF of each voxel was divided by the global mean ALFF value for each subject.

Regional Homogeneity (ReHo)

Based on the hypothesis that intrinsic brain activity is manifest by clusters of voxels rather than single voxels, ReHo evaluates the degree of regional similarity or synchronization of fMRI time courses (Zang et al., 2004). It is defined as the Kendall's coefficient concordance (KCC) (Kendall and Gibbons, 1990) of time series within a given voxel and its nearest neighbors. In the current analyses, the number of neighboring voxels was set to 26, which included voxels on the faces, edges, and the corners of a given voxel. For standardization purpose, as used in the ALFF calculation, the ReHo of each voxel was divided by the global mean ReHo value for each subject.

Voxel-Mirrored Homotopic Connectivity (VMHC)

Functional homotopy, the synchrony in spontaneous activity between geometrically corresponding interhemispheric regions, is a fundamental characteristic of the brain's functional architecture (Salvador et al., 2005). It can be quantified by calculating the Pearson correlation coefficient between each voxel's time series and that of its symmetric inter-hemispheric counterpart (Zuo et al., 2010). Correlation values were then transformed by Fisher's Z-transformation ($z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right)$) to approach a normal distribution.

Functional Connectivity Strength (FCS)

FCS at a voxel is defined as the average functional connectivity (FC) between that given voxel and all other voxels in the brain, i.e., $FCS_i = \frac{1}{N-1} \sum_{j \neq i} FC_{ij}$ (Liang et al., 2013). In this experiment, we only considered FCS within gray matter (GM) voxels. Pearson correlation coefficients between each voxel and all other voxels in an individual's GM mask were calculated and transformed into z-scores using Fisher's Z-transformation; FCS maps were then computed. Here the GM mask was derived from the segmentation step during data preprocessing.

We used the Resting-State fMRI Data Analysis Toolkit (Song et al., 2011) to calculate the ALFF, ReHo, and VMHC maps. For all local measures, mean values were extracted from individuals using the 116 region standard Automated Anatomical Labeling

(AAL) template (Tzourio-Mazoyer et al., 2002) which served as input features for the classifiers. The use of the AAL template aimed to reduce feature dimensions and improve signal-to-noise ratio.

Network Measures

We categorize the following as network measures given that they characterize the relationship between pairs of brain regions (i.e., edges in graph theory). Two kinds of network measures were considered in this study: One was seed-based brain networks, measuring the correlation between one voxel cluster (i.e., the seed) and all other voxels in the brain; the other, including temporal correlation and Granger causality, was the interaction of signals between all pairs of AAL regions.

Seed-Based Brain Networks

Seed-based methods were applied using AFNI to extract five widely studied large-scale brain networks: the default-mode network (DMN), executive-control network (ECN), salience network (SN), striatum network (StrN), and limbic network (LN). Notably, the DMN, ECN, and SN have been implicated to work in an interacting fashion, including in nicotine dependence (Sridharan et al., 2008; Bressler and Menon, 2010; Bonnelle et al., 2012; Sutherland et al., 2012; Jilka et al., 2014; Lerman et al., 2014; Liang et al., 2015; Uddin, 2015). Additionally, the StrN and LN are two networks putatively related to drug addiction (Kelley and Berridge, 2002; David et al., 2005; Everitt and Robbins, 2005; Gu et al., 2010; Janes et al., 2012). Seed regions were defined by placing bilateral 3 mm radius spherical regions of interest (ROIs) in the posterior cingulate cortex (PCC) as an exemplar constituent of the DMN (Greicius et al., 2003), the dorsal lateral prefrontal cortex (dlPFC) for the ECN (Seeley et al., 2007), the insula for the SN (Seeley et al., 2007), the caudate for the StrN (Di Martino et al., 2008), and the amygdala for the LN (Gu et al., 2010); see **Table 2** for center coordinates of seeds. For each brain network seed, a reference time course was generated by averaging the time course from all voxels within the ROI. Subsequently, a correlation coefficient (CC) map was obtained by correlating each voxel's time course with the corresponding reference time course. The CC maps were then transformed by Fisher's Z-transformation into z-score maps. Finally, these z-scored brain network maps were partitioned into 116 ROIs using the standard AAL atlas, and mean values within each AAL region served as input features for the classifiers.

TABLE 2 | Seeds locations used to define brain networks.

Brain networks	Seed location	Center (Talairach coordinates)
Default-mode network (DMN)	PCC	(±2, -51, 27)
Executive-control network (ECN)	dlPFC	(±43, 36, 21)
Salience network (SN)	Insula	(±40, 6, -6)
Limbic network (LN)	Amygdala	(±23, -5, -15)
Striatum network (StrN)	Caudate	(±8, 6, -4)

Temporal Correlation (TC)

Functional connectivity refers to the functionally integrated relationship between different brain regions regardless of the apparent physical connectedness (Friston, 2011). One definition is the TC between spatially remote neurophysiological events (Biswal et al., 1995). In contrast to the above seed-based network measures, which compute the functional connectivity between a well-defined, *a priori* hypothesized ROI and all other voxels, the TC here is calculated pairwise using mean time series extracted from standard AAL template regions. Pearson correlation coefficients were computed between mean time series extracted from the 116 standard AAL regions, and transformed into z-scores using Fisher's Z-transformation. Due to symmetry, we only took the lower triangle z-score matrices as our input features.

Granger Causality (GC)

In contrast to temporal correlation, one may attempt to measure causal influence exerted by one neuronal system onto another (Goebel et al., 2003; Friston, 2011). Granger causality analysis (GCA) has been proposed to estimate the causal interactions of information flow. It models one directional causality among multiple time series based on a vector autoregression (VAR) model (Seth, 2005). When the model's residual error reaches the minimum, an *F*-test is used to estimate the statistical significance of the estimated model. A higher *F*-score means a stronger prediction of GC between two time series. We employed a Matlab toolbox for GCA (Seth, 2010) to calculate the GC between mean time series of uncensored data within 116 regions standard AAL atlas. The VAR model order was estimated using Akaike information criterion (AIC) (Burnham and Anderson, 2004). Resulting *F*-score matrices were treated as input features. It should be noted that the use of GCA applied to neuroimaging data is controversial (Friston et al., 2013). However, we make no claims of its ability to determine any causal relationship between regions; it is merely another feature that may convey complementary information to improve group discrimination classification accuracy (Deshpande et al., 2010).

In this study, we used the AAL template, which is an anatomical atlas based template, to define our ROIs from which features were extracted from each resting-state feature type. Another common way to process resting state data is to use independent components analysis (ICA) to define brain networks. Some studies extracted ICA components from the respective groups to conduct machine learning (Van Waarde et al., 2015) or resting state functional connectivity (Cerliani et al., 2015) analysis. In an effort to compare the AAL template with ICA-derived regions, we additionally applied a network-based ROI template approach where networks were generated from publicly available group ICA maps (Smith et al., 2009). Details are described in Supplementary Materials.

Feature Selection Using Grid-Search

In contrast to studies that selected discriminative features lower than a statistical threshold (Fan et al., 2007; Deshpande et al., 2010; Dai et al., 2012; Feis et al., 2013; Hart et al., 2014), we determined the optimal percentage of reserved features using a

grid-search method for the three frameworks described below. Specifically, for each feature type, with the aid of an inner 10-fold CV, features were first sorted based on their T -scores using a two-sample t -test (Pereira et al., 2009; Chu et al., 2012; Mwangi et al., 2014). An RBF-SVM was then used to search for an optimized feature size within percentage values (1, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50%) of the sorted features. Thus, in the inner fold, the optimized feature size varied for different feature types and for different folds, which means that the function of performance along with the size of reserved features changed in each inner fold for each feature type. The LIBSVM toolbox (Chang and Lin, 2011) was used for all classification procedures. Two hyperparameters including the regularization constant C and Gaussian kernel parameter γ in the RBF-SVM were optimized using a nested 10-fold CV among the values of 2^N (N from -4 to 6 for C and from -10 to 3 for γ). In addition to the optimized feature size, we also recorded its corresponding accuracy in the inner 10-fold CV for later use.

Three Optimized Frameworks

To address the combinatorial options of the multiple feature types, we implemented three optimized frameworks: feature combination, kernel combination, and classifier combination.

Feature Combination Framework

The feature combination framework illustrated in **Figure 1** performed a multi-feature type combination before classifier training. In this framework, selected features from each type were concatenated into a row vector and input to an RBF-SVM classifier. Hyperparameters C and γ in the RBF-SVM were optimized using a nested 10-fold CV among the values of 2^N (N from -4 to 6 for C and from -10 to 3 for γ).

Kernel Combination Framework

As mentioned in the Introduction, features are more likely to be linearly separable when they are projected into a higher dimensional space through a kernel-induced implicit mapping function. A well-known property of kernels is that they can be combined via linear operations to yield a new valid kernel. Let $\mathbf{x}_i^{(m)}$ denote a feature vector in the m -th feature type of the i -th sample whose class label is $y_i \in \{-1, 1\}$. Multi-kernel SVM aims to solve the following primal problem:

$$\begin{aligned} \min_{\mathbf{w}^{(m)}, b, \varepsilon_i} & \frac{1}{2} \sum_{m=1}^M \beta_m \|\mathbf{w}^{(m)}\|^2 + C \sum_{i=1}^n \varepsilon_i \\ \text{s.t. } & y_i \left(\sum_{m=1}^M \beta_m \left((\mathbf{w}^{(m)})^T \mathbf{K}^{(m)}(\mathbf{x}_i^{(m)}) + b \right) \right) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, \beta_m \geq 0, i = 1 \dots n \end{aligned} \tag{1}$$

Here, $\mathbf{w}^{(m)}$, $\mathbf{K}^{(m)}$, β_m , ε_i , C , and b denote, respectively, the weight vector of hyperplane, the kernel-induced mapping function, the combining weight on the kernel, the non-negative slack variable, the trade-off, and the offset of the hyperplane. A more detailed description of multi-kernel SVM can be found in (Gonen and Alpaydin, 2011). In this experiment, selected features from each type were projected into a higher dimensional space using RBF as the kernel mapping function.

Since the main idea of multi-kernel SVM is to first construct an individual kernel for each feature type and then train a mixed kernel based on the linear combination of all individual kernels (Zhang D. et al., 2011), similar to Gonen and Alpaydin (2011); Zhang D. et al. (2011) and Zhang and Shen (2012), we added a constraint $\sum_{m=1}^M \beta_m = 1$ to the kernel combining weights. Considering the large number of feature types (11 in total) in our experiment, and in contrast to prior work utilizing this technique (Zhang D. et al., 2011; Zhang and Shen, 2012) that applied a coarse grid-search method to determine the combining weights for only three modalities, we chose a heuristic approach in which the recorded accuracy of each feature type from the above feature selection procedure was used to choose the combining weights in the following form:

$$\begin{cases} \beta_m = \frac{Acc_m - 0.5}{\sum_{i=1}^M (Acc_i - 0.5)}; & \text{if } Acc_m > 0.5, \text{ and for all } Acc_i > 0.5 \\ \beta_m = 0; & \text{if } Acc_m \leq 0.5 \end{cases} \tag{2}$$

where Acc_m denotes the accuracy of m -th feature type via inner 10-fold CV in the feature selection. The other hyperparameters in the multi-kernel SVM were optimized using the above described grid-search method. The kernel combination framework is illustrated in **Figure 2**.

Classifier Combination Framework

Discriminative information from multiple feature types can also be combined after classifier training, which is the basis of our classifier combination framework (see **Figure 3**). In this framework, selected features from each type were input to a RBF-SVM classifier. Hyperparameters were optimized as described above. Let $f_m(x_i)$ be an output decision value of the SVM classifier on m -th feature type for i -th sample, a final classifier was then combined using a weighted voting approach:

$$y_i = \text{sgn} \left(\sum_{m=1}^M \beta_m f_m(x_i) \right), \text{ s.t., } \sum_{m=1}^M \beta_m = 1 \tag{3}$$

Here, β_m is the classifier combining weight for m -th feature type, and was determined using the same weighting scheme described above.

Cross-Validation

As illustrated in **Figures 1–3**, all three frameworks were evaluated using a balanced outer 10-fold CV procedure. That is, in each outer trial, 10 smokers and 10 non-smokers were excluded before feature selection (i.e., they were left out of the whole analysis) for testing the classifier that was trained using all other subjects. The classification quality was assessed by the following five quantities:

$$\text{Sensitivity} = TP / (TP + FN) \tag{4}$$

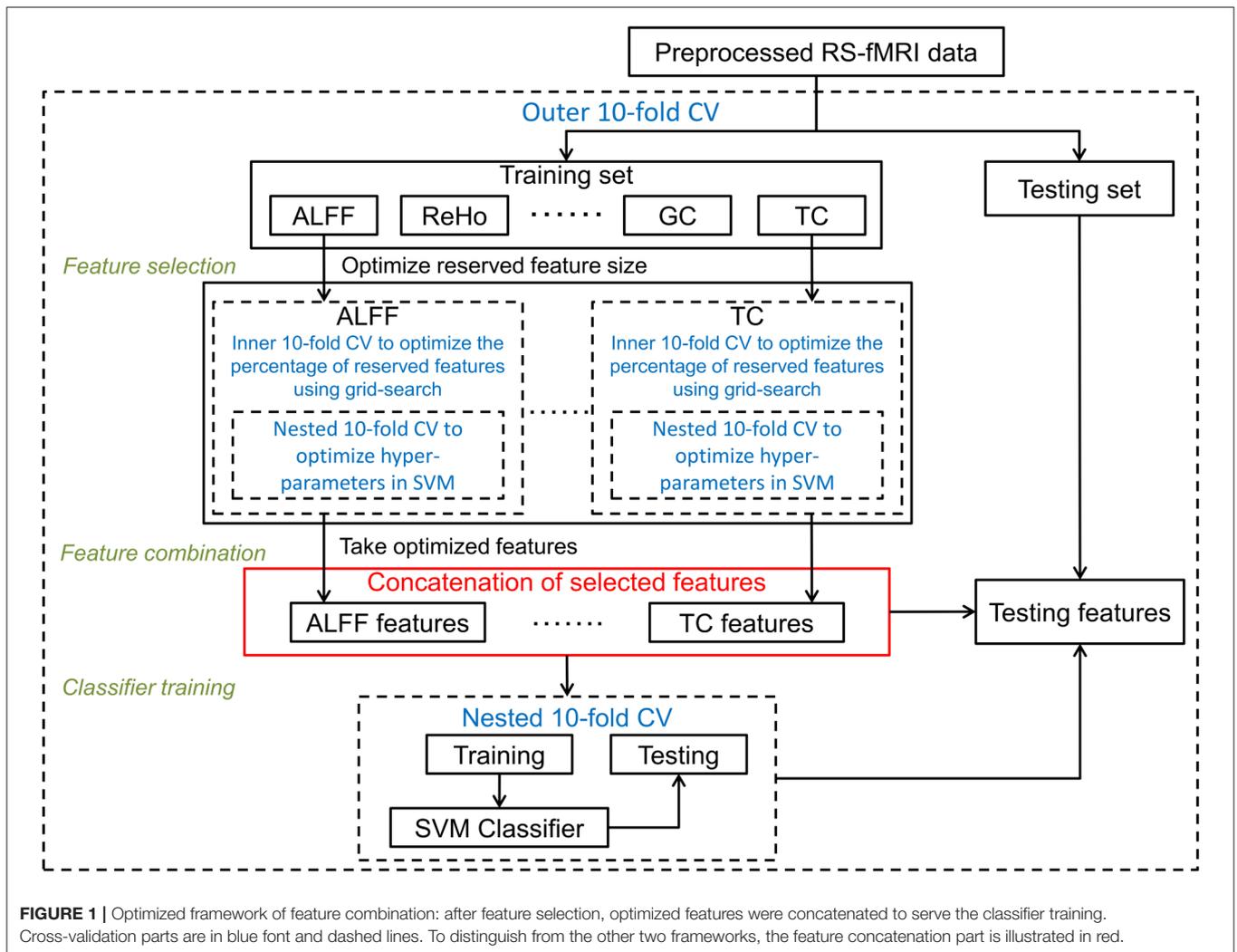
$$\text{Specificity} = TN / (TN + FP) \tag{5}$$

$$\text{Accuracy} = (TP + TN) / (TP + FN + TN + FP) \tag{6}$$

$$\text{Precision} = TP / (TP + FP) \tag{7}$$

$$F \text{ score} = 2TP / (2TP + FP + FN) \tag{8}$$

Here, TP, FN, TN, and FP denote, respectively, the number of smokers correctly classified, the number of smokers predicted to



be non-smokers, the number of non-smokers correctly classified, and the number of non-smokers predicted to be smokers. Specifically, sensitivity, also called the true positive rate, measures the proportion of smokers that are correctly identified as such; while specificity, also called the true negative rate, measures the proportion of non-smokers that are correctly identified as such. Precision, also called positive predictive value, is the proportion of smokers that are identified as such; and F score is the harmonic mean of precision and sensitivity. We did not employ the receiver operating characteristic (ROC) calculation to assess our frameworks; since in our classifier combination framework, a SVM classifier was trained on each feature set, it would be unreasonable to move the cut-off thresholds in the same range for each SVM to plot a ROC curve.

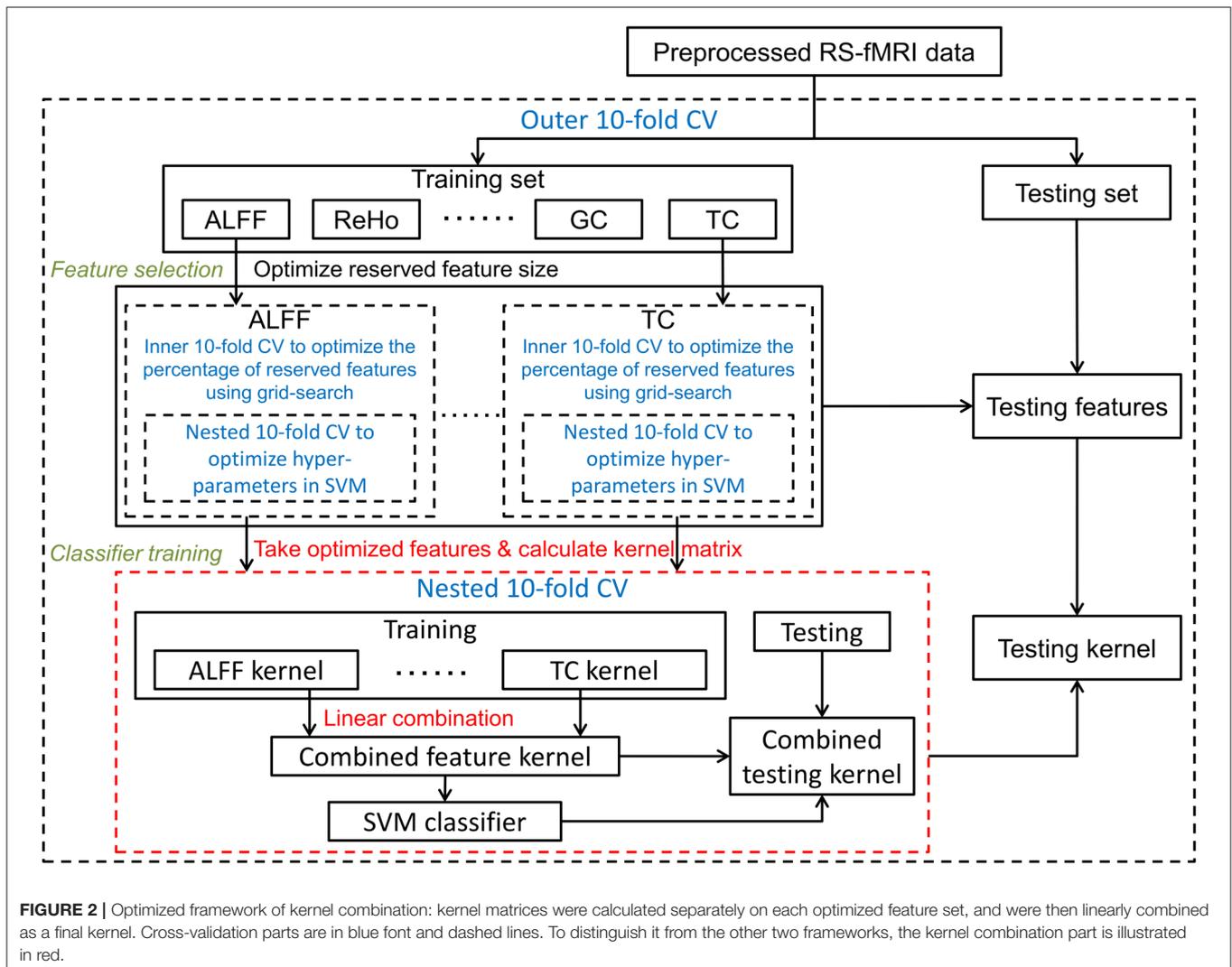
Finally, we performed significance analysis on the selected feature maps: For each resting-state feature type in feature selection, we recorded the percentage of features that was reserved in each fold from the three frameworks (30-folds in total). To determine the threshold of significant ROIs for each feature type, we first randomly chose ROIs according to the

recorded percentage (i.e., if 5% of the regions were retained for that fold, we randomly choose 5% of the ROIs to assess significance). We then calculated the number of times that an ROI was randomly selected among all 30-folds. This whole process was repeated 1,000 times to derive an empirical null distribution. The actual data were thresholded at $P < 0.05$ based upon the empirical null.

RESULTS

Classification Performance

Classification results of the three tested frameworks are shown in **Table 3**. Using nicotine dependence as a model system and using all feature types, the three approaches overall yielded very similar results; the classifier combination and the feature combination frameworks reached an accuracy of 75.5%, while the kernel combination achieved a 73.0% accuracy. As a comparison, the discriminative ability of each feature type was tested individually (see **Table 4**), where the RBF-SVM classifiers were performed separately on each type using the same feature selection and



hyperparameter optimizing framework. As a single feature type, TC achieved the highest accuracy of 70.5%, but most of the other feature types individually only performed at slightly above chance with the exception of GC. Notably, all proposed combination frameworks improved the classification accuracy over any single feature type. Since GC showed the lowest performance (accuracy = 49.0%) among the single feature types, we implemented the three combination frameworks following GC elimination. Somewhat paradoxically, the accuracy of all three frameworks slightly decreased (see **Table 3**) when excluding this feature type that performed worse than chance on its own, indicating that even the worst feature contributed some information to the combination frameworks.

Compared to the AAL template, all frameworks performed worse when using the ICA-based template (see Supplementary Materials for details). Given the superior classifier accuracy and brain coverage given by the AAL template, we report only those results in this manuscript.

A box plot of β_m , used in the kernel combination and classifier combination frameworks, from Equation (2) shows that the

feature type TC consistently had the highest weight, while other feature types performed similarly (see **Figure 4**). Also illustrated is that the GC was consistently the worst feature and was frequently weighted zero.

Maps of Significant ROIs

The threshold to reach significance for a given ROI ($P < 0.05$) was determined to be 9-folds for the ALFF, 15-folds for the FCS, 15-folds for the ReHo, 17-folds for the VMHC, 14-folds for the DMN, 10-folds for the ECN, 12-folds for the LN, 16-folds for the SN, 15-folds for the StrN, 17-folds for the TC, and 22-folds for the GC. Significant feature maps are shown in **Figure 5**. ROIs in the prefrontal cortex (PFC), subcortical regions (e.g., the thalamus, caudate, putamen), occipital lobe, and cerebellum were significant in the maps of many of the feature types. Among these, the thalamus was significant in both local (e.g., ALFF, ReHo, VMHC) and network measures (e.g., DMN, LN). Additionally, the TC between the subcortical regions and the frontal cortex as well as the cerebellum significantly differentiated smokers

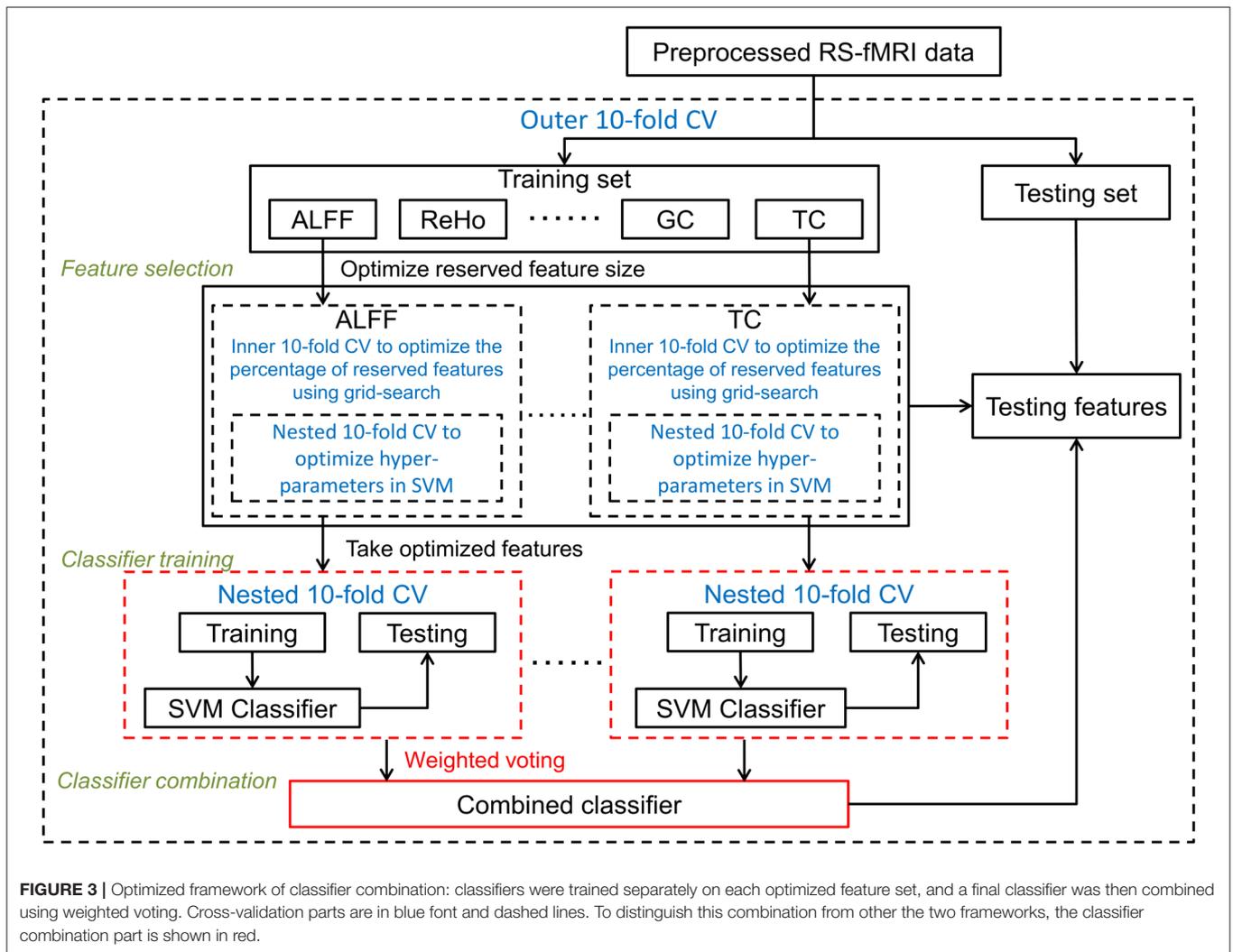


TABLE 3 | Classification results of the three tested frameworks.

		Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F score (%)
Feature combination	All features	76.0	75.0	75.5	75.2	75.6
	Without GC	73.0	71.0	72.0	71.6	72.3
Kernel combination	All features	77.0	69.0	73.0	71.3	74.0
	Without GC	74.0	68.0	71.0	69.8	71.8
Classifier combination	All features	79.0	72.0	75.5	73.8	76.3
	Without GC	74.0	75.0	74.5	74.7	74.4

from non-smokers, suggesting abnormal functional connectivity between these regions in smokers.

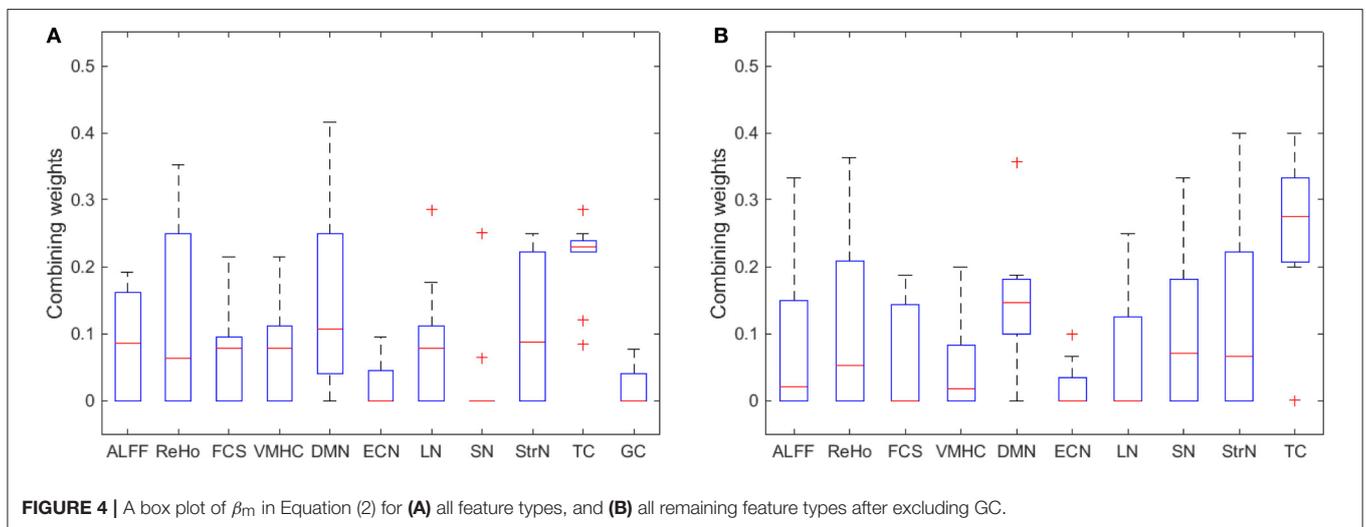
DISCUSSION

In recent years, machine learning techniques have become widely applied to neuroimaging data to predict neurological and psychiatric disorders (Orrù et al., 2012; Wolfers et al., 2015; Arbabshirani et al., 2016), with the long-term goal to create

complex brain-based biomarkers of disease status that could enormously benefit the treatment community. The scientific motivation of this study was to examine various techniques to combine multiple resting-state feature types so as to utilize their complementary information to maximize classification accuracy, thus this is first and foremost a methods paper. We proposed three frameworks addressing SVM classification using multi-type features: feature combination, kernel combination, and classifier combination. We chose nicotine addiction as an exemplar disease

TABLE 4 | Discriminative ability of individual feature types.

		Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F score (%)
Local measures	ALFF	60.0	53.0	56.5	56.1	58.0
	ReHo	59.0	61.0	60.0	60.2	59.6
	FCS	46.0	67.0	56.5	58.2	51.4
	VMHC	46.0	66.0	56.0	57.5	51.1
Network measures	DMN	62.0	64.0	63.0	63.3	62.6
	ECN	58.0	42.0	50.0	50.0	53.7
	LN	66.0	41.0	53.5	52.8	58.7
	SN	54.0	57.0	55.5	55.7	54.8
	StrN	61.0	54.0	57.5	57.0	58.9
	TC	76.0	65.0	70.5	68.5	72.0
	GC	48.0	50.0	49.0	49.0	48.5

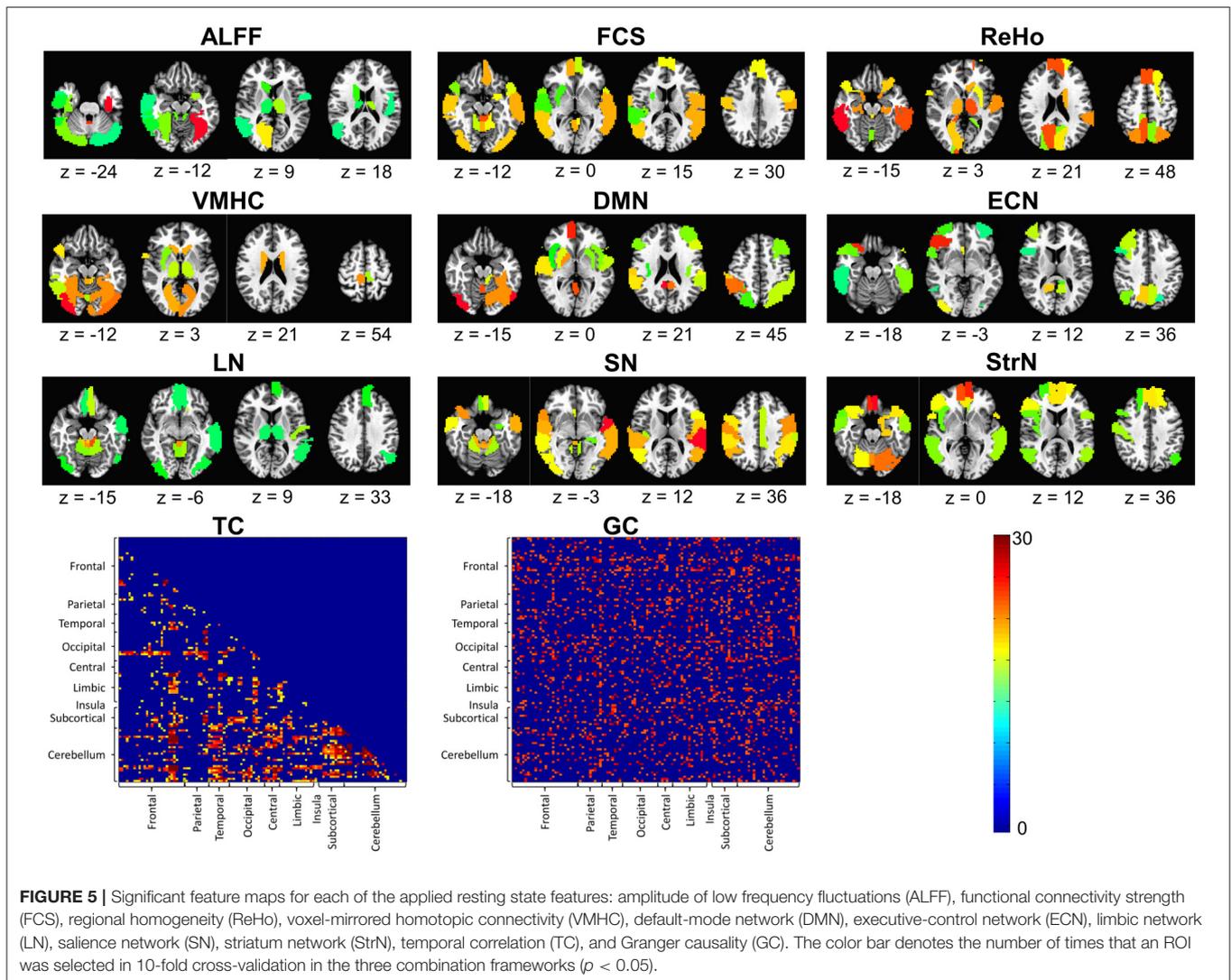
**FIGURE 4** | A box plot of β_m in Equation (2) for **(A)** all feature types, and **(B)** all remaining feature types after excluding GC.

model system and tested the frameworks on 11 resting-state features consisting of both local and network measures to predict smoking status. All frameworks were validated using a 10-fold CV procedure, and all demonstrated an improvement over the classification performance using any one of the 11 single feature types.

We designed a grid search approach involving two-sample *t*-tests to sort the features for selection. Feature selection approaches are classified into “filter,” “wrapper,” and “embedded” methods (Pereira et al., 2009; Mwangi et al., 2014). The approach of the two-sample *t*-test that we chose falls under the “filter” category. Although some studies argue that *t*-test filtering is not stable and robust enough since it is only performed once using the training data (Venkataraman et al., 2010), it has been demonstrated that combining *t*-test filtering and atlas based ROI leads to significantly better accuracy than no feature selection when sample sizes are small (Chu et al., 2012). Another well-known available feature selection method is recursive feature elimination (RFE), which is a “wrapper” method. We chose, however, not to use it as it would be prohibitively computationally time consuming for our nested CV

design with the number of features included herein because in RFE, features are sorted and the least discriminative feature is eliminated. This procedure is repeated iteratively until all features are tested.

Previously, (Pettersson-Yeo et al., 2014) also chose SVM as an elementary classifier on raw feature sets calculated from multiple modalities of imaging data to do multi-feature combination. In contrast to their method that combined raw feature maps to train classifiers without any feature selection, we designed an optimized feature selection procedure using a grid-search method on AAL atlas based ROI features ranked by two-sample *t*-tests. Advantages of using the AAL atlas are that it’s anatomically based and the most commonly used ROI template; however, a potential limitation is that there is no direct physiological relationship between AAL regions and neurobiological processing units or nicotine addiction, making data interpretation neurobiologically difficult. In our hands, using smoking as a model neuropsychiatric disease (Leshner, 1997; Hasin et al., 2013), we observed improved classification using ROI based features over the raw voxel wise features (Ding



et al., 2015), as the spatial averaging improves signal-to-noise. Using ROIs also benefited the classifier training in the combination frameworks, as it reduced feature dimension and improved the grid-search speed for eliminating less informative features.

Another major difference from Pettersson-Yeo et al. (2014) lies in the combination frameworks employed. Multi-kernel learning algorithms can be divided into one-step methods or two-step methods in terms of their training methodology (Gonen and Alpaydin, 2011). One-step methods using fixed rules and heuristics generally do not have much computational complexity to determine the kernel combination weight, whereas two-step methods update the combination weights by solving an optimization problem whose convergence may be slow or even hard to get for a large number of kernels. Thus, in the kernel combination framework, rather than using either an unweighted simple sum of kernels approach or a complex two-step learning algorithm (Pettersson-Yeo et al., 2014), we chose a heuristic approach using the accuracy of each feature type

in the feature selection procedure as kernel combining weights, a simplification necessary with a large number of kernels. In the classifier combination framework, considering that the SVM classifiers trained using different feature types had different performances, we used a weighted voting approach instead of a simple prediction averaging or majority voting method (Pettersson-Yeo et al., 2014).

A previous work by our group applied a SVM-based classification procedure to rsFC data from 21 smokers and 21 non-smokers. That work mainly focused on testing different characteristics of nicotine dependence related network connectivity to predict smoking status. The classifier achieved an accuracy of 78.6% using within-network functional connectivity measures via LOOCV (Pariyadath et al., 2014). In contrast, the present work focused on methods for combining multiple resting-state feature types using different classification frameworks. Moreover, it was generated from a larger dataset of 100 smokers and 100 non-smokers, which would have been expected to result in a more reliable classification result. Besides

the difference in sample size, the difference in classification accuracy may also have been caused by different cross-validation procedures (i.e., LOOCV vs. 10-fold cross-validation). In particular, LOOCV, necessary in that study due to the modest sample size, is known to yield anticonservative (i.e., over-fitting) results (Kohavi, 1995). As such, we believe that the current work represents a better estimate of what is possible with rsFC in a smoking model system.

Subcortical brain areas and cerebellum, prominent in our feature maps, are thought to be involved in functional networks supporting higher-order executive function and top-down control, including in various addiction studies (Hester and Garavan, 2004; Dosenbach et al., 2008; Goldstein and Volkow, 2011). For example, the thalamus, shown as one of the most discriminative regions in many of our feature maps, has previously been shown related to nicotine addiction neurobiology (Rubboli et al., 1994; Stein et al., 1998; Franklin et al., 2007; Hahn et al., 2009; Beaver et al., 2011). Additionally, a key alpha5 nicotinic receptor gene variant is associated with a dorsal anterior cingulate-ventral striatum/extended amygdala circuit that distinguishes smokers from non-smokers and predicts addiction severity in smokers (Hong et al., 2010). Moreover, nicotine improves sustained attention by increasing activation in the thalamus, caudate, and occipital lobe (Lawrence et al., 2002). Additional discriminative regions identified in the present study are located in the prefrontal cortex (PFC), which are known to play a key role in addictive behaviors through regulation of limbic regions and its involvement in higher-order executive functions (Goldstein and Volkow, 2011; Zhang X. et al., 2011).

All of the proposed combination frameworks improved classification accuracy over using a single feature type, and are easily applied to cases of resting-state data classification problems. It is worth noting that other imaging data phenotypes (e.g., gray matter density) or even non-imaging measures (e.g., genetics, behavioral, or personality phenotypes) can be added into these frameworks as feature types. In our hands, the feature combination and the classifier combination framework performed slightly better than the kernel combination when using the AAL template; and the classifier combination framework performed better than the other two frameworks when using ICA generated ROIs, although critically, none of these differences would survive a statistical test. Nevertheless, the present results may usefully guide future studies and encourage the use of whichever method investigators are most familiar with or the simplest approach (i.e., feature combination). However, since we only applied these frameworks to a nicotine addiction case as an exemplar, it is not known if a given framework would clearly outperform the other two frameworks when applied to other classification cases, using other templates for feature extraction or indeed other types of classifier input data (e.g., anatomical measures). Nonetheless, our study is an important contribution to the literature as it informs others facing similar choices.

One limitation of this study is that, like many other methodological studies (for example, Demirci et al., 2008; Yang et al., 2010; Castro et al., 2011; Dai et al., 2012; Jie et al., 2015;

Kim et al., 2016), our feature combination methods were only validated using one disease exemplar. Notably, our experimental results on nicotine addiction fell into a moderate accuracy range (subject to neither ceiling nor floor effects), which is consistent with the extant literature using machine learning techniques to predict neuropsychiatric disorders (Orrù et al., 2012; Wolfers et al., 2015; Arbabshirani et al., 2016). Given these results, we believe that the proposed method would have good generalizability to other disease exemplars. Another important limitation of this study is that of a limited data size, as larger data sets are known to improve classifier accuracy. Future studies should address these issues.

CONCLUSION

In this study, we proposed three optimized frameworks: feature combination, kernel combination, and classifier combination, which we examined separately, to combine multiple types of resting-state features calculated from categories of local measures and network measures into classification. These frameworks were successfully applied to a nicotine dependence case, demonstrating their efficacy in improving classification performance over using a single feature type. Our proposed frameworks have good generalizability and can be applied to other neuropsychiatric diseases with extended feature types from other data phenotypes.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of NIH Human Research Protection Program (HRPP) policies and human subjects protections regulations with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Addictions Institutional Review Board of the National Institute on Drug Addiction and National Institute on Alcohol Abuse and Alcoholism.

AUTHOR CONTRIBUTIONS

XD and TR designed the study. XD conducted the analyses and drafted the manuscript. All authors gave contribution to the data collection, result interpretation, manuscript revision, and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institute on Drug Abuse and by FDA grant number NDA13001-001-00000 to EAS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2017.00362/full#supplementary-material>

REFERENCES

- Arbabshirani, M. R., Plis, S., Sui, J., and Calhoun, V. D. (2016). Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage* 145(Pt B), 137–165. doi: 10.1016/j.neuroimage.2016.02.079
- Beaver, J. D., Long, C. J., Cole, D. M., Durcan, M. J., Bannon, L. C., Mishra, R. G., et al. (2011). The effects of nicotine replacement on cognitive brain activity during smoking withdrawal studied with simultaneous fMRI/EEG. *Neuropsychopharmacology* 36, 1792–1800. doi: 10.1038/npp.2011.53
- Behzadi, Y., Restom, K., Liu, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101. doi: 10.1016/j.neuroimage.2007.04.042
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409
- Bonnelle, V., Ham, T. E., Leech, R., Kinnunen, K. M., Mehta, M. A., Greenwood, R. J., et al. (2012). Salience network integrity predicts default mode network function after traumatic brain injury. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4690–4695. doi: 10.1073/pnas.1113455109
- Bressler, S. L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290. doi: 10.1016/j.tics.2010.04.004
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167. doi: 10.1023/A:1009715923555
- Burnham, K., and Anderson, D. (2004). Multimodel inference - understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33, 261–304. doi: 10.1177/0049124104268644
- Castro, E., Martínez-Ramón, M., Pearlson, G., Sui, J., and Calhoun, V. D. (2011). Characterization of groups using composite kernels and multi-source fMRI analysis data: application to schizophrenia. *Neuroimage* 58, 526–536. doi: 10.1016/j.neuroimage.2011.06.044
- Cerliani, L., Mennes, M., Thomas, R. M., Di Martino, A., Thioux, M., and Keyers, C. (2015). Increased functional connectivity between subcortical and cortical resting-state networks in autism spectrum disorder. *JAMA Psychiatry* 72, 767–777. doi: 10.1001/jamapsychiatry.2015.0101
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 27. doi: 10.1145/1961189.1961199
- Chen, R., and Herskovits, E. H. (2010). Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage* 52, 234–244. doi: 10.1016/j.neuroimage.2010.03.084
- Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., Lin, C., and Initiative, A. S. D. N. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60, 59–70. doi: 10.1016/j.neuroimage.2011.11.066
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014
- Dai, Z., Yan, C., Wang, Z., Wang, J., Xia, M., Li, K., et al. (2012). Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage* 59, 2187–2195. doi: 10.1016/j.neuroimage.2011.10.003
- David, S. P., Munafò, M. R., Johansen-Berg, H., Smith, S. M., Rogers, R. D., Matthews, P. M., et al. (2005). Ventral striatum/nucleus accumbens activation to smoking-related pictorial cues in smokers and nonsmokers: a functional magnetic resonance imaging study. *Biol. Psychiatry* 58, 488–494. doi: 10.1016/j.biopsych.2005.04.028
- Demirci, O., Clark, V. P., and Calhoun, V. D. (2008). A projection pursuit algorithm to classify individuals using fMRI data: application to schizophrenia. *Neuroimage* 39, 1774–1782. doi: 10.1016/j.neuroimage.2007.10.012
- Deshpande, G., Li, Z., Santhanam, P., Coles, C. D., Lynch, M. E., Hamann, S., et al. (2010). Recursive cluster elimination based support vector machine for disease state prediction using resting state functional and effective brain connectivity. *PLoS ONE* 5:e14277. doi: 10.1371/journal.pone.0014277
- Di Martino, A., Scheres, A., Margulies, D. S., Kelly, A. M., Uddin, L. Q., Shehzad, Z., et al. (2008). Functional connectivity of human striatum: a resting state fMRI study. *Cereb. Cortex* 18, 2735–2747. doi: 10.1093/cercor/bhn041
- Ding, X., and Lee, S. W. (2013). Changes of functional and effective connectivity in smoking replenishment on deprived heavy smokers: a resting-state fMRI study. *PLoS ONE* 8:e59331. doi: 10.1371/journal.pone.0059331
- Ding, X., Yang, Y., Stein, E. A., and Ross, T. J. (2015). Multivariate classification of smokers and nonsmokers using SVM-RFE on structural MRI images. *Hum. Brain Mapp.* 36, 4869–4879. doi: 10.1002/hbm.22956
- Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., and Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn. Sci.* 12, 99–105. doi: 10.1016/j.tics.2008.01.001
- Eloyan, A., Muschelli, J., Nebel, M. B., Liu, H., Han, F., Zhao, T., et al. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front. Syst. Neurosci.* 6:61. doi: 10.3389/fnsys.2012.00061
- Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi: 10.1038/nn1579
- Fan, Y., Rao, H., Hurt, H., Giannetta, J., Korczykowski, M., Shera, D., et al. (2007). Multivariate examination of brain abnormality using both structural and functional MRI. *Neuroimage* 36, 1189–1199. doi: 10.1016/j.neuroimage.2007.04.009
- Fedota, J. R., and Stein, E. A. (2015). Resting-state functional connectivity and nicotine addiction: prospects for biomarker development. *Ann. N. Y. Acad. Sci.* 1349, 64–82. doi: 10.1111/nyas.12882
- Feis, D. L., Brodersen, K. H., Von Cramon, D. Y., Luders, E., and Tittgemeyer, M. (2013). Decoding gender dimorphism of the human brain using multimodal anatomical and diffusion MRI data. *Neuroimage* 70, 250–257. doi: 10.1016/j.neuroimage.2012.12.068
- Franklin, T. R., Wang, Z., Wang, J., Sciortino, N., Harper, D., Li, Y., et al. (2007). Limbic activation to cigarette smoking cues independent of nicotine withdrawal: a perfusion fMRI study. *Neuropsychopharmacology* 32, 2301–2309. doi: 10.1038/sj.npp.1301371
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Friston, K., Moran, R., and Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* 23, 172–178. doi: 10.1016/j.conb.2012.11.010
- Gabrieli, J. D., Ghosh, S. S., and Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85, 11–26. doi: 10.1016/j.neuron.2014.10.047
- Goebel, R., Roebroeck, A., Kim, D. S., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21, 1251–1261. doi: 10.1016/j.mri.2003.08.026
- Goldstein, R. Z., and Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat. Rev. Neurosci.* 12, 652–669. doi: 10.1038/nrn3119
- Gonen, M., and Alpaydin, E. (2011). Multiple kernel learning algorithms. *J. Mach. Learn. Res.* 12, 2211–2268.
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100
- Gu, H., Salmeron, B. J., Ross, T. J., Geng, X., Zhan, W., Stein, E. A., et al. (2010). Mesocorticolimbic circuits are impaired in chronic cocaine users as demonstrated by resting-state functional connectivity. *Neuroimage* 53, 593–601. doi: 10.1016/j.neuroimage.2010.06.066
- Hahn, B., Ross, T. J., Wolkenberg, F. A., Shakleya, D. M., Huestis, M. A., and Stein, E. A. (2009). Performance effects of nicotine during selective attention, divided attention, and simple stimulus detection: an fMRI study. *Cereb. Cortex* 19, 1990–2000. doi: 10.1093/cercor/bhn226
- Hart, H., Chantiluke, K., Cubillo, A. I., Smith, A. B., Simmons, A., Brammer, M. J., et al. (2014). Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. *Hum. Brain Mapp.* 35, 3083–3094. doi: 10.1002/hbm.22386

- Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., et al. (2013). DSM-5 criteria for substance use disorders: recommendations and rationale. *Am. J. Psychiatry* 170, 834–851. doi: 10.1176/appi.ajp.2013.12060782
- Hester, R., and Garavan, H. (2004). Executive dysfunction in cocaine addiction: evidence for discordant frontal, cingulate, and cerebellar activity. *J. Neurosci.* 24, 11017–11022. doi: 10.1523/JNEUROSCI.3321-04.2004
- Hirsch, R. (1991). Validation samples. *Biometrics* 47, 1193–1194.
- Hong, L. E., Hodgkinson, C. A., Yang, Y., Sampath, H., Ross, T. J., Buchholz, B., et al. (2010). A genetically modulated, intrinsic cingulate circuit supports human nicotine addiction. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13509–13514. doi: 10.1073/pnas.1004745107
- Idaka, T. (2015). Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex* 63, 55–67. doi: 10.1016/j.cortex.2014.08.011
- Janes, A. C., Nickerson, L. D., Frederick, B. E. B., and Kaufman, M. J. (2012). Prefrontal and limbic resting state brain network functional connectivity differs between nicotine-dependent smokers and non-smoking controls. *Drug Alcohol Depend.* 125, 252–259. doi: 10.1016/j.drugalcdep.2012.02.020
- Jie, B., Zhang, D., Cheng, B., Shen, D., and Initiative, A. S. D. N. (2015). Manifold regularized multitask feature learning for multimodality disease classification. *Hum. Brain Mapp.* 36, 489–507. doi: 10.1002/hbm.22642
- Jilka, S. R., Scott, G., Ham, T., Pickering, A., Bonnelle, V., Braga, R. M., et al. (2014). Damage to the salience network and interactions with the default mode network. *J. Neurosci.* 34, 10798–10807. doi: 10.1523/JNEUROSCI.0518-14.2014
- Kelley, A. E., and Berridge, K. C. (2002). The neuroscience of natural rewards: relevance to addictive drugs. *J. Neurosci.* 22, 3306–3311.
- Kendall, M. G., and Gibbons, J. D. (1990). *Rank Correlation Methods*. London: Edward Arnold.
- Kim, J., Calhoun, V. D., Shim, E., and Lee, J. H. (2016). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage* 124, 127–146. doi: 10.1016/j.neuroimage.2015.05.018
- Klöppel, S., Abdulkadir, A., Jack, C. R., Koutsouleris, N., Mourão-Miranda, J., and Vemuri, P. (2012). Diagnostic neuroimaging across diseases. *Neuroimage* 61, 457–463. doi: 10.1016/j.neuroimage.2011.11.002
- Kohavi, R. (1995). “A study of cross-validation and bootstrap for accuracy estimation and model selection” in *International Joint Conference on Artificial Intelligence* (Montreal: IJCAI), 1137–1143.
- Lawrence, N. S., Ross, T. J., and Stein, E. A. (2002). Cognitive mechanisms of nicotine on visual attention. *Neuron* 36, 539–548. doi: 10.1016/S0896-6273(02)01004-8
- Lee, M. H., Smyser, C. D., and Shimony, J. S. (2013). Resting-state fMRI: a review of methods and clinical applications. *AJNR Am. J. Neuroradiol.* 34, 1866–1872. doi: 10.3174/ajnr.A3263
- Lerman, C., Gu, H., Loughhead, J., Ruparel, K., Yang, Y., and Stein, E. A. (2014). Large-scale brain network coupling predicts acute nicotine abstinence effects on craving and cognitive function. *JAMA Psychiatry* 71, 523–530. doi: 10.1001/jamapsychiatry.2013.4091
- Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Science* 278, 45–47. doi: 10.1126/science.278.5335.45
- Liang, X., Zou, Q., He, Y., and Yang, Y. (2013). Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1929–1934. doi: 10.1073/pnas.1214900110
- Liang, X., Zou, Q., He, Y., and Yang, Y. (2015). Topologically Reorganized connectivity architecture of default-mode, executive-control, and salience networks across working memory task loads. *Cereb. Cortex* 26, 1501–1511. doi: 10.1093/cercor/bhu316
- Liberio, L. E., Deramus, T. P., Lahti, A. C., Deshpande, G., and Kana, R. K. (2015). Multimodal neuroimaging based classification of autism spectrum disorder using anatomical, neurochemical, and white matter correlates. *Cortex* 66, 46–59. doi: 10.1016/j.cortex.2015.02.008
- Liu, F., Guo, W., Fouche, J. P., Wang, Y., Wang, W., Ding, J., et al. (2015). Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct. Funct.* 220, 101–115. doi: 10.1007/s00429-013-0641-4
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and Initiative, A. S. D. N. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002
- Mwani, B., Tian, T. S., and Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12, 229–244. doi: 10.1007/s12021-013-9204-3
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* 36, 1140–1152. doi: 10.1016/j.neubiorev.2012.01.004
- Pariyath, V., Stein, E. A., and Ross, T. J. (2014). Machine learning classification of resting state functional connectivity predicts smoking status. *Front. Hum. Neurosci.* 8:425. doi: 10.3389/fnhum.2014.00425
- Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007
- Pettersson-Yeo, W., Benetti, S., Marquand, A. F., Joles, R., Catani, M., Williams, S. C., et al. (2014). An empirical comparison of different approaches for combining multimodal neuroimaging data with support vector machine. *Front. Neurosci.* 8:189. doi: 10.3389/fnins.2014.00189
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048
- Power, J. D., Schlaggar, B. L., and Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105, 536–551. doi: 10.1016/j.neuroimage.2014.10.044
- Rehme, A. K., Volz, L. J., Feis, D. L., Bomilcar-Focke, I., Liebig, T., Eickhoff, S. B., et al. (2015). Identifying neuroimaging markers of motor disability in acute stroke by machine learning techniques. *Cereb. Cortex* 25, 3046–3056. doi: 10.1093/cercor/bhu100
- Richiardi, J., Achard, S., Bunke, H., and Van De Ville, D. (2013). Machine learning with brain graphs. *IEEE Signal Process. Mag.* 30, 58–70. doi: 10.1109/MSP.2012.2233865
- Rubboli, F., Court, J. A., Sala, C., Morris, C., Perry, E., and Clementi, F. (1994). Distribution of neuronal nicotinic receptor subunits in human brain. *Neurochem. Int.* 25, 69–71. doi: 10.1016/0197-0186(94)90055-8
- Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., and Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332–1342. doi: 10.1093/cercor/bhi016
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007
- Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network* 16, 35–54. doi: 10.1080/09548980500238756
- Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* 186, 262–273. doi: 10.1016/j.jneumeth.2009.11.020
- Shen, H., Wang, L., Liu, Y., and Hu, D. (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *Neuroimage* 49, 3110–3121. doi: 10.1016/j.neuroimage.2009.11.011
- Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045. doi: 10.1073/pnas.0905267106
- Song, X. W., Dong, Z. Y., Long, X. Y., Li, S. F., Zuo, X. N., Zhu, C. Z., et al. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE* 6:e25031. doi: 10.1371/journal.pone.0025031
- Sridharan, D., Levitin, D. J., and Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and

- default-mode networks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12569–12574. doi: 10.1073/pnas.0800005105
- Stein, E. A., Pankiewicz, J., Harsch, H. H., Cho, J. K., Fuller, S. A., Hoffmann, R. G., et al. (1998). Nicotine-induced limbic cortical activation in the human brain: a functional MRI study. *Am. J. Psychiatry* 155, 1009–1015. doi: 10.1176/ajp.155.8.1009
- Suk, H. I., Lee, S. W., and Shen, D. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct. Funct.* 220, 841–859. doi: 10.1007/s00429-013-0687-3
- Sundermann, B., Herr, D., Schwindt, W., and Pfeleiderer, B. (2014). Multivariate classification of blood oxygen level-dependent fMRI data with diagnostic intention: a clinical perspective. *AJNR Am. J. Neuroradiol.* 35, 848–855. doi: 10.3174/ajnr.A3713
- Sutherland, M. T., Mchugh, M. J., Pariyadath, V., and Stein, E. A. (2012). Resting state functional connectivity in addiction: lessons learned and a road ahead. *Neuroimage* 62, 2281–2295. doi: 10.1016/j.neuroimage.2012.01.117
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.* 16, 55–61. doi: 10.1038/nrn3857
- Van Waarde, J. A., Scholte, H. S., Van Oudheusden, L. J., Verwey, B., Denys, D., and 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. doi: 10.1038/mp.2014.78
- Venkataraman, A., Kubicki, M., Westin, C. F., and Golland, P. (2010). Robust feature selection in resting-state fMRI connectivity based on population studies. *Conf. Comput. Vis. Pattern Recognit. Workshops* 2010, 63–70. doi: 10.1109/cvprw.2010.5543446
- Wang, Y., Fan, Y., Bhatt, P., and Davatzikos, C. (2010). High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. *Neuroimage* 50, 1519–1535. doi: 10.1016/j.neuroimage.2009.12.092
- Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., and Marquand, A. F. (2015). From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* 57, 328–349. doi: 10.1016/j.neubiorev.2015.08.001
- Wu, G., Yang, S., Zhu, L., and Lin, F. (2015). Altered spontaneous brain activity in heavy smokers revealed by regional homogeneity. *Psychopharmacology* 232, 2481–2489. doi: 10.1007/s00213-015-3881-6
- Yang, H., Liu, J., Sui, J., Pearlson, G., and Calhoun, V. D. (2010). A hybrid machine learning method for fusing fmri and genetic data: combining both improves classification of Schizophrenia. *Front. Hum. Neurosci.* 4:192. doi: 10.3389/fnhum.2010.00192
- Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., et al. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* 29, 83–91. doi: 10.1016/j.braindev.2006.07.002
- Zang, Y., Jiang, T., Lu, Y., He, Y., and Tian, L. (2004). Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22, 394–400. doi: 10.1016/j.neuroimage.2003.12.030
- Zeng, L. L., Shen, H., Liu, L., and Hu, D. (2014). Unsupervised classification of major depression using functional connectivity MRI. *Hum. Brain Mapp.* 35, 1630–1641. doi: 10.1002/hbm.22278
- Zhang, D., and Shen, D. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* 59, 895–907. doi: 10.1016/j.neuroimage.2011.09.069
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zhang, X., Salmeron, B. J., Ross, T. J., Geng, X., Yang, Y., and Stein, E. A. (2011). Factors underlying prefrontal and insula structural alterations in smokers. *Neuroimage* 54, 42–48. doi: 10.1016/j.neuroimage.2010.08.008
- Zuo, X. N., Kelly, C., Di Martino, A., Mennes, M., Margulies, D. S., Bangaru, S., et al. (2010). Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. *J. Neurosci.* 30, 15034–15043. doi: 10.1523/JNEUROSCI.2612-10.2010

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ding, Yang, Stein and Ross. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.