



Multi-modal multiple kernel learning for accurate identification of Tourette syndrome children



Hongwei Wen^{a,b,c,1}, Yue Liu^{e,1}, Islem Rekik^g, Shengpei Wang^{a,b,c}, Zhiqiang Chen^{a,b,c},
Jishui Zhang^f, Yue Zhang^e, Yun Peng^{e,*}, Huiguang He^{a,b,c,d,**}

^a State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

^b Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China

^c University of Chinese Academy of Sciences, Beijing, China

^d Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing, China

^e Department of Radiology, Beijing Children's Hospital, Capital Medical University, Beijing, China

^f Department of Neurology, Beijing Children's Hospital, Capital Medical University, Beijing, China

^g Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

ARTICLE INFO

Keywords:

Tourette syndrome
DTI
TBSS
SVM
MKL

ABSTRACT

Tourette syndrome (TS) is a childhood-onset neurobehavioral disorder characterized by the presence of multiple motor and vocal tics. To date, TS diagnosis remains somewhat limited and studies using advanced diagnostic methods are of great importance. In this paper, we introduce an automatic classification framework for accurate identification of TS children based on multi-modal and multi-type features, which is robust and easy to implement. We present in detail the feature extraction, feature selection, and classifier training methods. In addition, in order to exploit complementary information revealed by different feature modalities, we integrate multi-modal image features using multiple kernel learning (MKL). The performance of our framework has been validated in classifying 44 TS children and 48 age- and gender-matched healthy children. When combining features using MKL, the classification accuracy reached 94.24% using nested cross-validation. Most discriminative brain regions were mostly located in the cortico-basal ganglia, frontal cortico-cortical circuits, which are thought to be highly related to TS pathology. These results show that our method is reliable for early TS diagnosis, and promising for prognosis and treatment outcome.

1. Introduction

Tourette syndrome (TS) is a common, chronic neuropsychiatric disorder characterized by the presence of multiple motor and vocal tics. The typical age of onset ranges from 5 to 7 years of age, and the majority of children improve by their late teens or early adulthood [1]. TS is frequently concomitant with obsessive-compulsive disorder (OCD), attention-deficit-hyperactivity disorder (ADHD) and other social and behavioral disturbances [2]. It is also typically diagnosed by observing symptoms and by evaluating the history of their onset. So far, only clinical measures have been developed to diagnose TS, while overlooking the informative power of brain morphological features. Doctors mostly need to carry out additional examination to eliminate other etiological conditions [1]. While tics constitute the major

diagnostic symptom, TS is quite heterogeneous and has a long-term prognosis that is difficult to accurately estimate. To date, TS diagnosis remains somewhat limited and studies using advanced diagnostic methods are still lacking.

With the progress of neuroimaging, many studies focused on identifying brain alterations associated with TS patients, which can potentially be used as a biomarker of TS pathophysiologic pattern. Accumulated evidence showed that both brain functional and structural changes are associated with TS. In particular, the dysfunction of the cortico-striatal-thalamo-cortical (CSTC) circuits was implicated in the pathophysiology of TS [3–7]. However, these findings are mainly obtained based on group-level statistical comparison, and thus are of limited value for individual-based disease diagnosis. To overcome this limitation, machine-learning techniques have been used in recent

Abbreviations: TS, Tourette syndrome; DTI, diffusion tensor imaging; WM, white matter; SVM, support vector machine; MKL, multiple kernel learning; TBSS, Tract-Based Spatial Statistics

* Corresponding author.

** Corresponding author at: State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

E-mail addresses: ppengyun@yahoo.com (Y. Peng), huiguang.he@ia.ac.cn (H. He).

¹ These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.patcog.2016.09.039>

Received 19 January 2016; Received in revised form 26 August 2016; Accepted 21 September 2016

Available online 22 September 2016

0031-3203/ © 2016 Elsevier Ltd. All rights reserved.

years, and have shown great potential in neuroimaging studies. Unlike group-based comparison approaches, machine-learning techniques are able to detect the fine-grained spatial discriminative patterns, which are critical for individual-based disease diagnosis.

Notably, these advanced methods have been applied to a range of MRI modalities in an effort to automate the diagnosis of different diseases such as autism spectrum disorder and mild cognitive impairment [8–10]. However, these studies were only based on one imaging modality, i.e., mostly on the structural MRI or DTI. Moreover, as indicated in the previous study [11], different biomarkers may provide complementary information. Therefore combining multi-modal features, instead of depending on one is a promising direction for improving classification accuracy. Intuitively, one can combine multiple results from different classifiers using a voting technique. Dai et al. [12] proposed a multi-classifier fusion model through weighted voting, using maximum uncertainty linear discriminant analysis (MLDA) as base classifiers, to distinguish Alzheimer's Disease (AD) patients from healthy controls. Moreover, combining several features into one single vector and then training a classifier can also be a practical option. Martin Dyrba et al. [13] used the same method to combine DTI and MRI features for the automated AD detection. In addition to the aforementioned fusion approaches, some novel methods were based on multi-task learning and multiple kernel learning (MKL) approaches. Zhang et al. [14] proposed a generic Multi-Modal Multi-Task (M3T) learning, to jointly predict multiple variables which can be used for classification, and the method was validated on multi-modal MRI in ADNI dataset, achieving better performance for classification tasks than various conventional learning methods. Dai et al. [15] developed a method that integrated multi-modal image features using MKL and achieved an excellent ranking in the ADHD-200 global competition, which is a world-wide classification contest on the ADHD-200 datasets. Although so many efforts have been recently made to investigate the potential of combining neuroimaging and machine-learning techniques for diagnostic purposes, few study was related to TS [16].

In our study, we aimed to take a critical step toward this goal and predict subject-specific TS diagnosis using structure MRI (sMRI) and DTI. Information from structural T1-weighted magnetic resonance imaging (MRI) can be used to evaluate the volume and density of cerebral white matter (WM) tissue and gray matter (GM) structures in order to find disease-related atrophy. Diffusion tensor imaging (DTI) can be used to assess the integrity of cerebral WM fiber tracts and, hence, can support the diagnosis of TS. From DTI, we can derive scalar indices of anisotropic diffusion, with the most widely used being the four metrics: fractional anisotropy (FA), reflecting the degree of directionality within the fiber tracts; radial diffusivity (RD), expression of diffusivity perpendicular to the fibers; axial diffusivity (AD), expression of diffusivity parallel to the fibers; and mean diffusivity (MD), representing the free water diffusion. Reduced FA [17,18] or increased RD [19,20] indicate impaired WM fiber tract integrity. Higher MD values generally indicate tissue damage, whereas changes in AD values have been hypothesized to differentiate respectively axonal injury from demyelination in WM tracts [21]. Specifically, analyzing the volumetric T1 scans and diffusion tensor imaging (DTI) of TS patients, using voxel-wise methods, such as voxel-based morphometry (VBM) [22] and tract-based spatial statistics (TBSS) [23], can better delineate the changes occurring in the GM and WM, respectively [24]. While VBM analysis is able to demonstrate regional GM atrophy, DTI is sensitive to microscopic changes in white matter (WM) integrity, which is not always detectable with standard sMRI [25]. So we only used VBM method with sMRI to extract GM volume features. Meanwhile, we used TBSS method with DTI to extract the FA, MD, AD and RD values as features.

Based on this work, the present study aims at combining sMRI data with DTI data in order to improve TS detection accuracy. We conducted extensive experiments on the selection of features, which were extracted from sMRI and DTI, as well as the combination of

multi-modal features using the multiple kernel SVM (MK-SVM) [26]. Furthermore, we also identified the key discriminative brain regions for classification, which may be potential biomarkers for TS diagnosis and help reveal the pathological mechanism of TS.

2. Materials and methods

2.1. Subjects

TS patients were recruited from outpatient clinics in Beijing Children's Hospital from July 2012 to May 2015 (age: 8.98 ± 3.114 years, range: 3–16 years; 11 female). All patients met DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, text revision) criteria for TS. We also included 48 age and gender matched health controls in our study (age: 11.00 ± 3.495 years; range: 3–17 years; 17 female). We used a clinical interview and the Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS) [27] to diagnose OCD and used the German short version of Wender Utah rating scale (WURS-k, translated to Chinese) [28] to diagnose ADHD. Patients fulfilling OCD criteria or other co-morbidities were excluded in the study. Tic severity for all patients was rated using the Yale Global Tic Severity Scale (YGTSS) [29] and ranged from 10 to 79 ([mean \pm SD]: 46.50 ± 18.037). The duration of TS ranged from 3 months to 5 years ([mean \pm SD]: 1.81 ± 1.423 years). For those who had course less than 1 year, TS diagnosis was made by follow-up call. After the study was approved by Beijing Children's Hospital review board, written informed consent was obtained from all the parents/guardians according to the Declaration of Helsinki. Details of the patients are shown Table 1.

2.2. Data acquisition

Magnetic resonance imaging was acquired using a 3.0 T MR scanner (Gyroscan Interna Nova, Philips, Netherland). Head positioning was standardized using canthomeatal landmarks. The head was stabilized with foam pads to minimize head movements. Patients were instructed to suppress tics and minimize head movements during scanning as much as possible. Axial three-dimensional diffusion tensor imaging (DTI) was acquired from all the subjects. DTI was performed using the following protocol: spin-echo diffusion-weighted echo-planar imaging sequence, 2 mm slice thickness, no inter-slice gap, repetition time=4300 ms, echo time=95 ms, field of view (FOV)= 255×255 mm, reconstructed image matrix= 336×336 . Diffusion MRI images were obtained from 30 non-collinear directions with a b value of 1000 s/mm². 3D T1-weighted imaging were performed with axial three-dimensional-Fast Field Echo (3D FFE) sequence with the following parameters: repetition time (TR)=25 ms, echo time (TE)=4.6 ms, flip angle=30°, reconstructed image matrix= 256×256 , field of view (FOV)= 200×200 mm, slice thickness =1 mm.

2.3. Overview of our method

The overview of our proposed classification pipeline is illustrated in Fig. 1. From the preprocessed DTI and structural MRI images, we first extracted the respective regional features. Further, a hybrid feature selection method which combines filter-based and wrapper-based methods was applied to select relevant features in each modality. Based on the selected features, a kernel matrix was constructed for each modality. Then, we fused the kernel individual matrices from each modality for SVM-based classification.

2.4. Data preprocessing

2.4.1. Structural MRI

2.4.1.1. Preprocessing using VBM-DARTEL. After the data acquisition, T1-weighted structural images were analyzed using

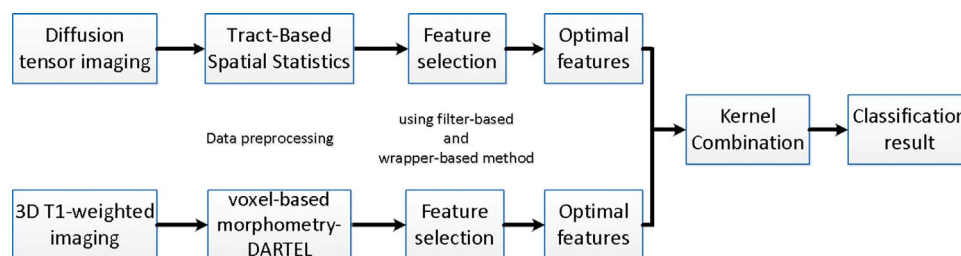
Table 1

Demographic variables and clinical characteristics of patients. Abbreviations: +=present, -=absent; YGTSS=Yale Global Tic Severity Scale; M=male; F=female.

ID	Sex	Age (y)	YGTSS	Duration	ADHD	OCD	Motor tics		Vocal tics	
							Simple	Complex	Simple	Complex
1	M	5	36	0.17	No	No	+	+	–	–
2	F	11	52	4	No	No	+	+	+	–
3	M	14	65	1.5	Yes	No	+	+	+	+
4	M	6	23	0.25	Yes	No	+	+	–	–
5	M	10	80	2	Yes	No	+	+	+	–
6	M	7	50	0.67	No	No	+	+	–	–
7	M	8	44	2	No	No	+	+	+	–
8	M	11	30	2	Yes	No	+	+	–	–
9	M	10	24	4	No	No	+	+	–	–
10	M	11	25	1	No	No	+	+	–	–
11	F	5	52	0.25	Yes	No	+	+	+	–
12	M	11	30	1.2	No	No	+	+	–	–
13	M	9	62	2.5	No	No	+	+	+	–
14	F	6	59	1	No	No	+	+	+	–
15	F	9	36	5.5	No	No	+	+	–	–
16	M	9	85	2	No	No	+	+	+	–
17	F	8	10	0.625	No	No	+	+	–	–
18	M	12	57	1	No	No	+	+	+	–
19	M	3	38	0.17	No	No	+	+	+	–
20	M	9	75	2	No	No	+	+	+	–
21	M	4	58	1	No	No	+	+	+	–
22	M	5	61	0.083	No	No	+	+	+	–
23	F	8	42	3	No	No	+	+	+	–
24	F	6	84	0.375	No	No	+	+	+	–
25	M	10	30	3	No	No	+	+	+	–
26	M	6	52	0.17	No	No	+	+	+	–
27	M	12	79	6	No	No	+	+	+	+
28	F	14	75	3	No	No	+	+	–	–
29	M	12	51	2	No	No	+	+	+	–
30	M	11	24	1	No	No	+	+	–	–
31	M	9	55	1.75	No	No	+	+	+	–
32	F	8	44	2	No	No	+	+	–	–
33	M	7	77	0.375	No	No	+	+	+	–
34	F	13	76	1.5	No	No	+	+	+	–
35	M	4	35	0.42	Yes	No	+	+	–	–
36	M	9	35	1	No	No	+	+	+	–
37	M	8	67	0.25	No	No	+	+	+	–
38	M	11	32	0.917	No	No	+	+	–	–
39	M	9	38	2	No	No	+	+	–	–
40	M	7	59	2.3	No	No	+	+	+	–
41	M	6	42	2.5	Yes	No	+	+	+	–
42	F	4	66	0.25	No	No	+	+	+	–
43	M	16	79	3.5	No	No	+	+	+	–
44	M	13	59	4	No	No	+	+	–	–

SPM8 package (Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, London, UK, <http://www.fil.ion.ucl.ac.uk/spm>) and the VBM protocol with modulation. We implemented the Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL) algorithm [30]. All the T1-weighted images were brain extracted to exclude the non-brain tissues and reoriented with the origin set close to the anterior commissure (AC). Then each reorientated image was segmented into GM, WM and CSF in native space using 'New Segment' SPM tool, and aligned GM and WM images were generated using a rigid transformation. The study-specific GM/WM templates were then created using the aligned images from all the

patients and controls using 'Run DARTEL (Create Template)' SPM tool. The procedure begins with the averaging of aligned data to generate an original template. Then, in the first step, for each subject, we warped their images to the estimated template. In the second step, using the warped images, we updated the template. These steps were iterated six times to obtain the final refined template, which was the average of the DARTEL registered data. During the template creation processing, all aligned images were warped to the template yielding a series of flow fields which parameterized the deformation in order to use in modulation to preserve actual GM and WM volume. Finally, transforming all the normalized, modulated data into MNI

**Fig. 1.** Schematic diagram illustrating the proposed classification framework.

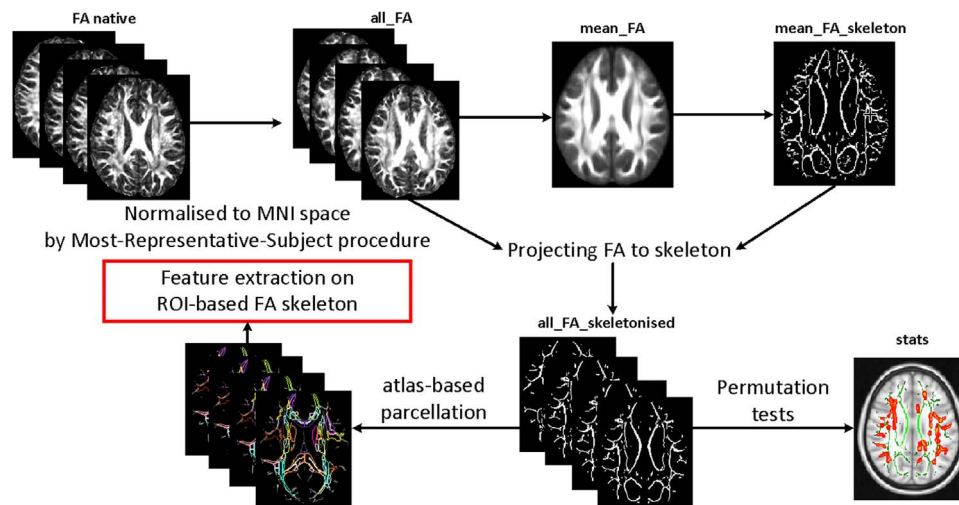


Fig. 2. Schematic diagram of Tract-Based Spatial Statistics analysis and feature extraction using atlas-based parcellation in our work. We use FA map as an example, the same projection achieved with FA images were applied to the MD, AD and RD maps.

space and smoothing all the images with an 8-mm FWHM isotropic Gaussian kernel were accomplished using 'Normalize to MNI Space' SPM tool. After the preprocessing, we obtained smoothed modulated normalized data (in the MNI space) to be used for the statistical analysis.

2.4.1.2. Feature extraction using atlas-based parcellation. We performed atlas-based parcellation to extract feature vectors from the segmented, normalized MR images using an anatomically labeled brain atlases: the AAL atlas. The AAL atlas is a single-subject atlas based on the Montreal Neurological Institute (MNI) Colin27 T1 atlas. This MNI single-subject brain template was obtained from 27 high-resolution T1-weighted scans of a young male. Each acquisition was spatially normalized to the MNI305 average template using a linear nine-parameter transformation [31]. In each hemisphere, 45 ROIs were manually drawn every 2 mm on the axial slices of the MNI single-subject brain. In addition, AAL includes a cerebellar parcellation with 26 ROIs [32]. Finally, 116 ROIs were defined, including the cerebellum for the AAL atlas.

We obtained the mean GM probability (density) for each ROI, as calculated by modulation with the Jacobian, which we used as a feature vector.

2.4.2. Diffusion tensor imaging

2.4.2.1. Preprocessing using TBSS. Following DTI acquisition, we used the FMRIB's Diffusion Toolbox (FDT2.0) within FSL v4.1 (<http://www.fmrib.ox.ac.uk/fsl>) for DTI processing. For each participant, 30 DTI volumes with 1000 s/mm² b-value were first affinely registered to the b₀ volume for correction of eddy current distortion and simple head motion. Non-brain voxels were removed using Brain Extraction Tool (BET) of FSL; a fractional intensity threshold of 0.25 was selected, resulting in a brain-extracted 4D image and a binary brain mask for each subject. We then used the eddy-corrected 4D data and corresponding brain mask to fit the diffusion tensor model at each voxel by using the FDT. Eigenvalues of diffusion tensor matrix ($\lambda_1, \lambda_2, \lambda_3$) were obtained and maps of axial diffusivity ($AD=\lambda_1$), mean diffusivity ($MD=(\lambda_1+\lambda_2+\lambda_3)/3$), and fractional anisotropy (FA) were generated. Radial diffusivity (perpendicular eigenvalue, $\lambda_{23}=(\lambda_2+\lambda_3)/2$) was calculated by averaging λ_2 and λ_3 maps.

The Most-Representative-Subject TBSS (RS-TBSS) procedure [33] was then applied to the data, which is recommended if the subjects are all young children. FA images from each participant were co-registered to one another using both linear (FLIRT) and nonlinear registration (FNIRT) provided by FSL. Next, we select the subject that has the minimum mean deformation for non-linearly aligning it to all subjects as the "most representative" one, and then used it as the target image. This target image is then affine-aligned into MNI152 standard space, and every image is transformed into 1×1×1 mm MNI152 space by combining the nonlinear transform to the target FA image using the affine transform from that target to MNI152 space. The resulting standard-space FA images for each participant were averaged to create a mean FA image, which was then fed into an FA skeletonization program to create a skeleton image of WM tracts; a threshold of 0.2 was selected to define the border of white and gray matter. Then, each participant's local maximal FA intensity along the perpendicular direction of the WM tract was projected to the mean FA skeleton to carry out the voxel-wise statistics across participants. Finally, the non-linear warps and the skeleton projection achieved with FA images were applied to the MD, AD and RD maps to bring them into standard space. In standard TBSS analysis, data were then fed into voxel-wise statistics with group comparisons. The permutation tool "randomize" was used, with 10,000 permutations and a significant threshold for between-group differences of $p < 0.05$.

2.4.2.2. Feature extraction using atlas-based parcellation. We performed atlas-based parcellation to extract feature vectors from the FA, MD, AD and RD TBSS skeleton using the Johns Hopkins University MNI template Type III White Matter Parcellation Map (Type III WMPM) (Fig. 2). The white matter and gray matter ROIs (118 areas) in the Type III WMPM was previously constructed by Oishi et al., and the detailed definition of ROIs could be found in [34]. Each ROI label was overlapped with the corresponding TBSS skeleton in the MNI space to obtain the mean FA, MD, AD and RD values on the skeleton within each ROI, which we used as a feature vector.

2.5. Feature selection

The dimensionality of the original ROI-based features extracted from VBM and TBSS analysis is much higher than the number of samples, which cannot be directly used to train a classifier considering the overfitting problem and the computational complexity. Hence, we employ a hybrid feature selection method which combines filter-based

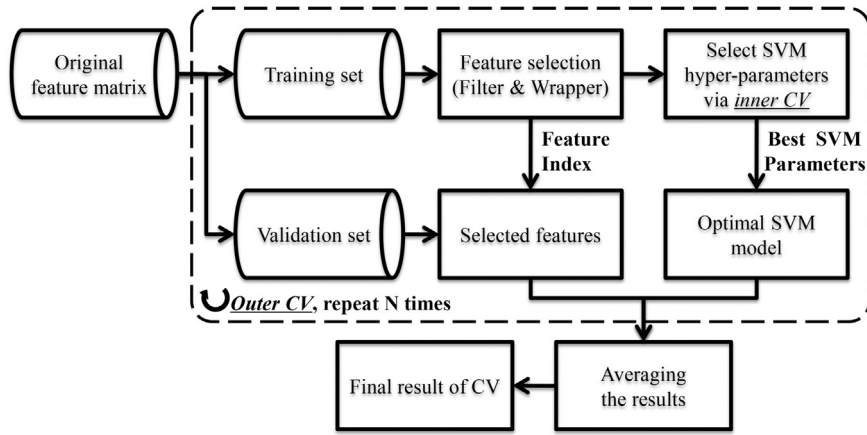


Fig. 3. The flow chart of the nested CV classification method.

and wrapper-based methods [35].

For filter-based method, we used the ratios of between-group sum of squares (BSS) to within-group sum of squares (WSS). The discriminative power of a feature can be estimated using BSS and WSS. A smaller WSS and a larger BSS usually mean that a feature is more prone to be distinguished. Thus, the ratio of BSS to WSS is used to rank features. Specifically, denoting f_{ij} as the value of the j -th feature of the i -th sample and r as the group label (+1 or -1), the ranking score of the j -th feature is defined as:

$$ratio_j = \frac{BSS_j}{WSS_j} = \frac{\sum_{i=1}^n \sum_{r=1}^2 I(y_i = r) (\bar{f}_{ij} - \bar{f}_j)^2}{\sum_{i=1}^n \sum_{r=1}^2 I(y_i = r) (f_{ij} - \bar{f}_{ij})^2} \quad (1)$$

where \bar{f}_{ij} is the mean value of the j -th feature across subjects in group r and \bar{f}_j is the mean value of the j -th feature across all n subjects; the index function $I(y_i=r)$ equals 1 if the i -th subject belongs to group r and equals 0 otherwise.

However, the aforementioned filter method computes the ranking scores independently for each feature, which does not take into account the relationship between features. So we employ a wrapper-based method for feature subset selection utilizing SVM based on recursive feature elimination, which is named SVM-RFE [36]. In this algorithm, SVM is trained iteratively using selected feature subset. In each iteration, the ranking score for each feature in the selected feature subset is calculated during SVM training process. The feature with the smallest score is eliminated in each iteration of SVM training until the classification accuracy is over a set threshold, or the number of remaining features in the selected subset is smaller than a set value. Note that SVM-RFE uses the accuracy of cross-validation (CV) to estimate the goodness of a feature subset, which may help avoid the overfitting problem. Thus, we first use the ratio of BSS/WSS (threshold of 0.05 in our study) to filter many features with little discriminative power, and then use SVM-RFE for further refining feature selection. This will ensure the selection of optimal feature subset at relatively low computational cost.

2.6. Classification methods

Classification was then performed using SVM, which is a supervised classification method that automatically learns a classification hyper-plane in the features space by optimizing a margin-based criterion. Let x_i represent a feature vector of the i -th subject and $y_i \in \{1, -1\}$ denote the corresponding class label. The primal optimization problem of the traditional SVM is given as:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \\ \text{s. t. } & y_i (w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

where ξ_i denotes non-negative slack variable which measures the degree of misclassification of the data, c denotes the penalty parameter which controls the amount of constraint violations introduced by ξ_i , b denotes the bias term, w denotes the normal vector of hyperplane and \cdot denotes the dot product.

We used an SVM with a radial basis function kernel to solve the classification problem. Let x_1, x_2 denote the feature vectors, and RBF kernel is defined as:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (3)$$

where σ is the width of the kernel. The hyper-parameters of SVM such as the penalized coefficient c that adjusts the importance of the separation error in the creation of the separation surface, and the kernel width σ should be carefully tuned to obtain the optimal SVM model.

In our framework, a nested 10-fold cross-validation strategy was used to evaluate the classification performance. For outer CV, the data is randomly divided into 10 parts in which each class is represented in approximately the same proportions as in the full dataset. Each fold is held out in turn and the learning scheme trained on the remaining nine-tenths and the error rate is then calculated on the tenth fold. Following 10 training procedures, we calculated the average CV accuracy and considered it as the estimation of generalization [37]. We performed another ten-fold cross-validation on the training samples, which is called inner CV. To estimate optimal values for c and σ we used a grid search in the range of $c=2^{-4}, 2^{-3}, \dots, 2^4$ and $\sigma=2^{-8}, 2^{-7}, \dots, 2^2$.

This nested CV method can yield an unbiased assessment of the classification method and prevent overestimation. Fig. 3 shows the flow chart of the evaluation method we used for nested CV.

Specifically, this nested CV procedure was repeated twenty times to avoid any bias introduced by randomly partitioning in the cross-validation.

2.7. Multiple kernel SVM

In addition to using SVM as a classifier, we also employed multi-kernel learning to integrate multi-modal features [26,38]. This method allows using different types of kernels or datasets concurrently in one single SVM instance, which works by simultaneously learning the predictor parameters and the kernel combination weights. The multiple kernels can come from different sources of feature spaces, thus providing a general framework for data fusion. It was previously used in successful applications to genomic data fusion, protein function prediction etc. Rakotomamonjy et al. [39] proposed the SimpleMKL algorithm, for solving the MKL problem through a weighted mixed-

norm regularization formulation with an additional constraint on the weights that encourages sparse kernel combinations. More recently, Vishwanathan et al. [40] proposed an SVM-based model to train p-norm Multiple Kernel Learning (MKL) and a more general linear MKL regularized by the Bregman divergence, using the Sequential Minimal Optimization (SMO) algorithm. Their algorithm retains both simplicity and efficiency and is significantly faster than state-of-the-art specialized p-norm MKL solvers.

The MLK algorithm can automatically search the optimal combination of the kernel matrix of these features to form an integrated kernel matrix (assigning weight to each kernel matrix), which outperforms the use of single kernels. The output of MKL is defined as:

$$y_i = \sum_k \beta_k \left(\sum_j \lambda_j^k y_j K_k(x_j^k, x_i^k) \right) + b \quad (4)$$

where k denotes the k -th kind of feature; y_i denotes the corresponding class label of i -th subject; K_k is the kernel matrix; β_k is the sub-kernel weight; λ_j^k is the Lagrange parameters; x_j^k is the support vector of training set; and x_i^k is the feature vector of the i -th test sample.

Although MKL is theoretically better, it has more parameters to tune, thus increasing the computational time for the training process. Hence, in our case, to reduce the computational load, we skip the feature selection and parameter tuning steps and directly use the same optimal feature subset and SVM parameters determined earlier for each single type of features. Our classification framework and validation experiments were implemented in Matlab using LIBSVM (v3.1.2, www.csie.ntu.edu.tw/~cjlin/libsvm/) for the SVM classifier and Shogun (v3.2.0, www.shogun-toolbox.org/) for the MKL framework.

2.8. Evaluation

For evaluation of results, we used accuracy, sensitivity, specificity, accuracy and the area under the curve for the receiver operated characteristic curve (AUC ROC). Accuracy is defined as $(TP+TN)/(TP+TN+FN+FP)$, sensitivity is defined as $TP/(TP+FN)$ and specificity is defined as $TN/(FP+TN)$, where TP is the number of true positives (number of patients correctly classified), TN is the number of true negatives (number of NC correctly classified), FP is the number of false positives (number of NC classified as patients), and FN is the number of false negatives (number of patients classified as NC). Besides, the area under ROC curve (AUC) denotes an evaluation measure derived from receiver operating characteristic (ROC) curve. The ROC curve is a graph evaluation method, which illustrates the performance of a binary classifier as its decision threshold is varied. When the decision threshold of a classifier varies, sensitivity and specificity also change. ROC curve is created by plotting sensitivity and 1-specificity at different thresholds. A larger AUC commonly indicates a better classifier.

3. Results

3.1. Classification results using a single feature

The SVM classifier is used, which classifies subjects into TS group and healthy control groups. The nested 10-fold CV classification results using a single feature are listed in Table 2, in which we compare the CV accuracies between five single-feature based classifiers which used ROI-based gray matter probability (GMP), ROI-based skeletonized FA, MD, AD and RD value as training features.

The bold font indicates the best performance using a single feature. The performance of MKL, which integrates all features, is listed in the bottom. Clearly, it produced higher classification accuracy than the methods using a single feature.

From the results listed in Fig. 4 and Table 2, we note that the classification using RD as a single feature achieves the best accuracy, sensitivity, specificity and the AUC. Fig. 5 shows the ROC curve of the

Table 2

SVM classification results using the single modalities and the multi-modal features.

Feature	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
GMP	76.30 ± 0.51	73.18 ± 6.41	79.17 ± 4.71	81.33 ± 3.90
FA	70.98 ± 0.31	70.68 ± 5.91	71.25 ± 5.96	77.36 ± 3.67
AD	76.41 ± 0.44	74.09 ± 4.44	78.54 ± 5.11	84.03 ± 2.96
RD	81.20 ± 0.46	79.77 ± 5.30	82.50 ± 5.12	88.62 ± 2.67
MD	70.00 ± 0.25	66.36 ± 3.18	73.33 ± 4.14	75.22 ± 3.69
MKL	94.24 ± 0.15	93.41 ± 3.46	95.00 ± 1.76	98.10 ± 0.95

CV classification using different features.

3.2. Classification based on combining features

To integrate the above five kinds of features, an MKL classifier is trained. When we combine all selected features, the accuracy reaches 94.24% using MKL-SVM method. This suggests that MKL that integrates multi-modal features is a powerful classifier, although it is more difficult to tune its hyper-parameters and is more time-consuming. The classification results using MKL are also listed in Table 2 and Fig. 4, and the ROC curve can be found in Fig. 5. We can see that the ROC curve of MKL classifier is better than single-feature based classifiers.

3.3. The most discriminative regions

In this subsection, we examine the most discriminative regions that were selected by the proposed feature selection method. Since the feature selection in each fold is performed only based on the training set, the selected features could differ slightly across different cross-validation folds. We thus define the most discriminative ROI-based features as features that were most frequently selected in all cross-validations. We summed the counts of each feature selected by our proposed method over the 20 rounds nested 10-fold CV. The top 22 selected features with a frequency exceeding 50% were provided in Table 2. The features of sMRI and DTI in the top 22 were shown in Figs. 6 and 7, respectively.

We also used two-sample t -test to evaluate the variation trend of features in the selected regions and whether each feature was significantly different in TS group compared to control group. In terms of DTI parameters, TS children showed significant decreased FA, significant increased RD and MD in all regions (see Table 3) compared to healthy controls. AD was significantly different in these regions between groups, but the variation trends were not consistent. As we can see, all variation trends of DTI parameters were consistent with previous studies as the introduction section mentioned. For gray matter probability (GMP), TS children showed significantly decreased GMP in cortical regions and significantly increased GMP in deep gray matter (basal ganglia).

It is worth noting that the frequency of AD feature in the right lateral fronto-orbital WM (LFOWM) and MD feature in the left superior cerebellar peduncle (SCP) exceeded 99%. Furthermore, multiple features in the lateral fronto-orbital WM (AD, FA in the right LFOWM and FA, RD in the left LFOWM) and the left SCP (MD, RD and AD) ranked within the top 22. This indicates that these regions may be highly related to TS pathology.

4. Discussion

We designed this study to investigate the macroscopic and microscopic changes in patients with TS pathology using MRI and to distinguish the TS children from normal children using a SVM classification based on imaging markers. We validated the proposed method on 92 subjects using a nested CV procedure, which considerably reduced the bias and gave an estimate of the error that is very close

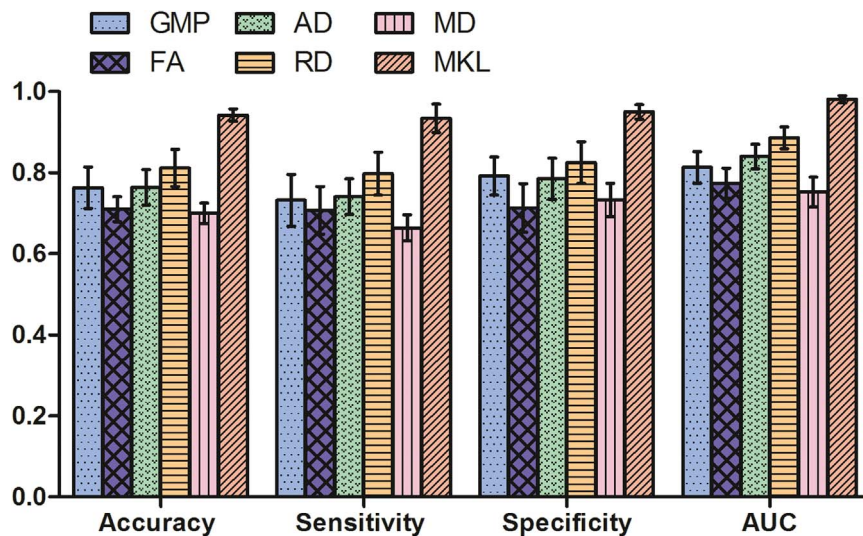


Fig. 4. Comparison of performances of different methods using the unimodal and multi-modal features.

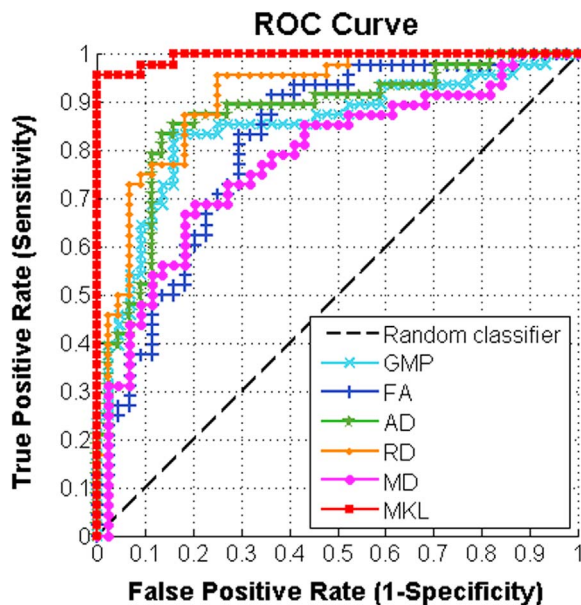


Fig. 5. The ROC curve of nested CV classification of 92 subjects.

to that obtained on the independent testing set. Our results spiked 94.24% of accuracy, which outperformed the previous work [16] that only extracted inter-hemispheric functional connectivity as features for TS classification, achieving a good classification accuracy (92.86%), sensitivity (91.67%), and specificity (93.75%). The detailed comparison between our proposed classification method against their method is listed below (Table 4).

4.1. Classification using multi-modal ROI-based features

Because of the difficulty of collecting samples, datasets in imaging studies are usually quite small in the neuroimaging community. Nearly one hundred samples with multi-modal MRI in our study represent a quite large dataset and cost thousands of dollars and several months or years to collect.

In our study, we used multi-modal MRI to investigate volumetric changes of GM using VBM and microstructural changes of WM using TBSS, and correlated these changes with identification of TS children. Compared with a single modality, the advantage of using multiple modalities is to extract more features (effective features) to robustly

learn a classifier. Theoretically, multi-modality method adopts different features as inputs, which can respectively reflect diverse aspects of samples. As previous structural and diffusion MR studies [24,41] showed widespread macroscopic (GMP) and microscopic (FA, AD, RD, MD) abnormalities in TS children, we believe that multi-modal and multi-type features can better reflect a wider spectrum of different classes' profiles and may lead to more accurate classification results. In fact, our results showed that the performance of multi-modal classifiers is better than the unimodal classifiers.

For classification, we used ROI-based features rather than voxel-based features since the use of ROI-based features reduces dimensionality and simplifies the clinical interpretation of pathology induced changes [42]. Of note, the ROI-based WM features were extracted using a novel atlas-based TBSS method, since our previous study showed that combining TBSS and atlas-based analysis can better reveal micro-structural abnormalities in early TS children [41]. Our findings also revealed some meaningful morphological and microstructural changes from structural MRI and DTI, which can be used as biomarkers for TS diagnosis.

4.2. Effect of hybrid feature selection method and MKL framework

In the current study, we proposed a multiple kernel learning framework for TS patients classification through combining multi-modal features. We extracted multi-modal features from the VBM and TBSS analysis and then used them to train classifiers. The hybrid feature selection algorithm, which combines filter and wrapper methods, was applied before training SVMs in order to prevent the overfitting problem and reduce the computational complexity. We also used a nested CV method to tune the hyper-parameters of classifiers and evaluate the performances of our method, which can yield unbiased estimation of classification method. In addition to using a single feature, we also employed MKL to integrate multi-modal features. Our experiments showed that MKL using multimodal features improved TS classification results when compared to unimodal classifiers.

4.3. Identify brain regions closely related to TS

Identification of objective biomarkers is of great interest as it could, ultimately, assist in clinical decisions for individual patients. With this consideration, our proposed feature selection method seeks to identify those ROI-based features that are most discriminative in distinguishing between TS children and controls, thereby identifying brain regions

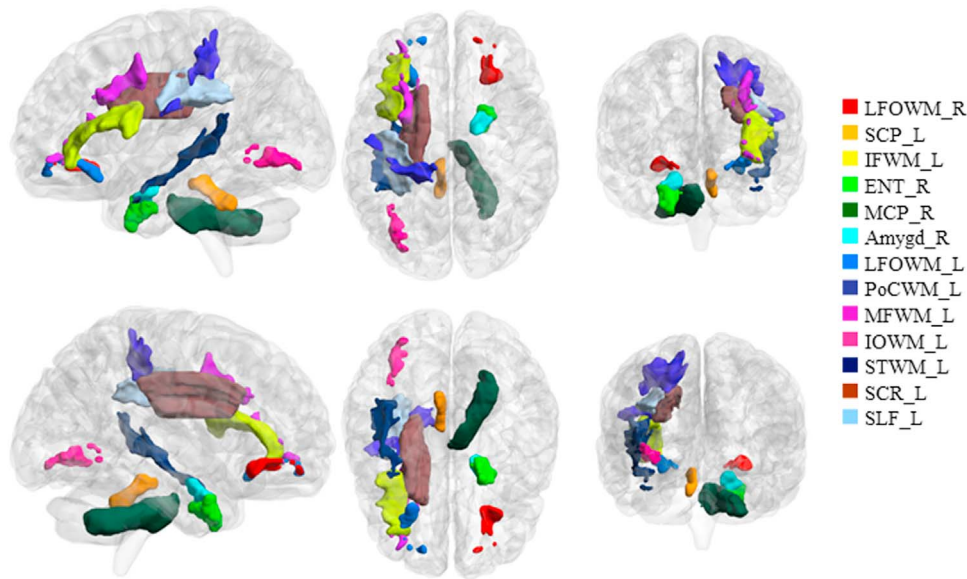


Fig. 6. Most discriminative regions derived from DTI for classifying TS and healthy children. Abbreviation is the same as Table 3.

that might be most related to TS.

The most discriminative regions in the TS classification were mainly located in superficial WM beneath frontal cortex, basal ganglia and deep WM fiber bundles.

4.3.1. Superficial white matter and related cortex

Many ROI-based features of frontal-related superficial white matter (SWM) ranked within the top 22, especially the right lateral fronto-orbital WM (LFOWM), with a high frequency of 99.5%. Moreover, multiple features in the bilateral LFOWM ranked within the top 22, indicating that these regions might be closely related to TS. The left inferior frontal WM (IFWM), which ranked third and left middle frontal WM (MFWM) was also frontal-related SWM.

Notably, SWM play an important role in highly and effectively control brain function [43]. In particular, SWM is WM lying just beneath the cortical ribbon. It is primarily composed of U-shaped association fibers which form the major local white matter connections in the brain arching through the cortical sulci to connect adjacent gyri, and intracortical axons which extend directly to white matter from the

overlying gray matter [44].

It is worth noting that the LFOWM lies beneath the lateral orbito-frontal cortex, where the lateral orbito-frontal circuit projecting from. The lateral orbito-frontal circuit connects to different parts of the caudate and globus pallidus, and projects to the thalamus and back to the orbito-frontal cortex. This circuit is thought to be involved in the control of inhibitory responses during learning and recognition tasks requiring frequent shifts of set developing during infancy. This may explain the perseveration or repetitive compulsive behavior seen with damage to the orbito-frontal cortex [45]. Of note, the lateral orbito-frontal circuit is one of five distinct parallel CSTC circuits that subserve different functions, and the interconnected relationships between movement disorders and associated behaviors may be understood by these parallel CSTC circuits [46]. Previous anatomical, functional, and lesion studies suggested that TS is caused by a failure of CSTC circuits to inhibit the somatosensory urges and associated motor enactments that constitute tic behaviors [47].

In addition to the lateral orbito-frontal circuit, the anterior cingulate circuit, which is also one of distinct parallel CSTC circuits

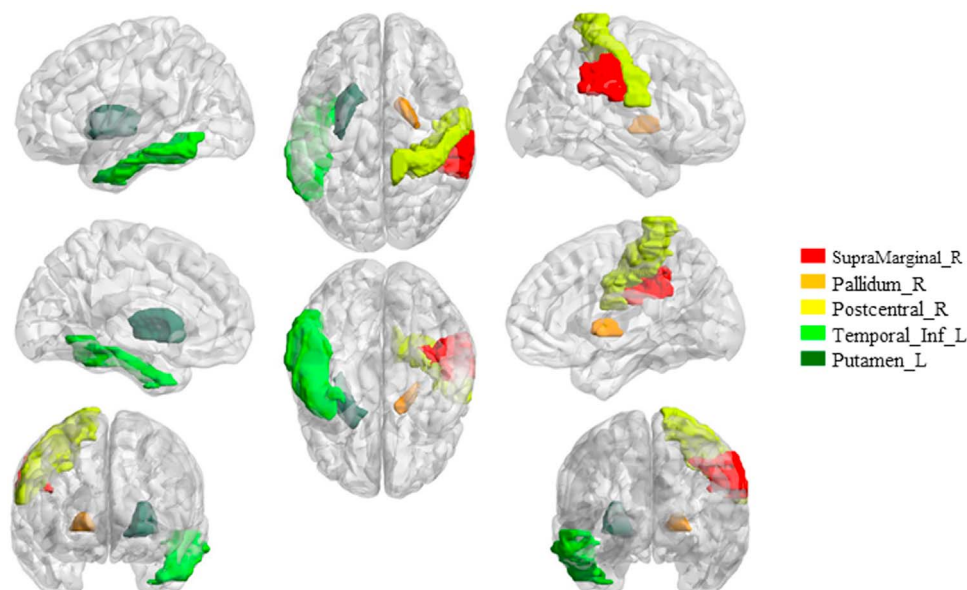


Fig. 7. Most discriminative regions derived from sMRI for classifying TS and healthy children. Abbreviations are the same as in Table 3.

Table 3

The top twenty-two ROI-based features of sMRI and DTI selected by the nested CV method as the most salient for group classification.

Region	Fea	Num	Rate	Trend	P	Abbreviation
Right lateral fronto-orbital WM	AD	199	99.50%	–	0.048	LFOWM_R
Left superior cerebellar peduncle	MD	198	99.00%	+	< 0.001	SCP_L
Left inferior frontal WM	FA	174	87.00%	–	< 0.001	IFWM_L
Right entorhinal area	AD	169	84.50%	–	0.017	ENT_R
Right middle cerebellar peduncle	RD	161	80.50%	+	0.001	MCP_R
Right Amygdala	AD	156	78.00%	+	0.049	Amygd_R
Right lateral fronto-orbital WM	FA	154	77.00%	–	< 0.001	LFOWM_R
Left superior cerebellar peduncle	RD	150	75.00%	+	< 0.001	SCP_L
Right supramarginal gyrus	GMP	150	75.00%	–	0.005	SupraMarginal_R
Left lateral fronto-orbital WM	FA	149	74.50%	–	0.001	LFOWM_L
Right globus pallidus	GMP	148	74.00%	+	0.049	Pallidum_R
Left postcentral WM	FA	143	71.50%	–	0.001	PoCWM_L
Right postcentral gyrus	GMP	136	68.00%	–	0.005	Postcentral_R
Left inferior temporal gyrus	GMP	130	65.00%	–	0.01	Temporal_Inf_L
Left putamen	GMP	129	64.50%	+	0.049	Putamen_L
Left middle frontal WM	FA	123	61.50%	–	< 0.001	MFWM_L
Left inferior occipital WM	FA	121	60.50%	–	0.001	IOWM_L
Left superior cerebellar peduncle	AD	121	60.50%	+	0.015	SCP_L
Left lateral fronto-orbital WM	RD	116	58.00%	+	0.002	LFOWM_L
Left superior temporal WM	AD	105	52.50%	+	0.006	STWM_L
Left superior corona radiata	AD	101	50.50%	+	0.016	SCR_L
Left superior longitudinal fasciculus	AD	101	50.50%	+	0.008	SLF_L

Num: the counts of each feature selected by our proposed method over the 20 rounds nested 10-fold CV. Rate: the frequency of being selected, equals Num/total times in 20 rounds nested 10-fold CV (200 times). ±: parameter is increased/decreased in TS group compared to control group. P: p value of two sample *t*-test. Fea: type of feature.

Table 4

The detailed comparison between our proposed classification method against the previous method [16] is listed below.

Paper	Liao et al. [16]	Present
Number of subjects	56	92
Feature Type	inter-hemispheric functional connectivity (one type)	GM volume, FA, AD, RD, MD (five types)
Feature selection	Multivariate Pattern Analysis	Hybrid method (filter-based method+SVM-RFE)
Classifier	SVM	SVM
Multi-type features fusion	No	Yes
Accuracy	92.86%	94.24%
Sensitivity	91.67%	93.41%
Specificity	93.75%	95.00%

[46], includes the ventral striatum, nucleus accumbens, and medial dorsal nucleus of the thalamus. The entorhinal area with the hippocampus are thought to send inputs to this circuit, which integrates information from the para-limbic association cortex [45]. Our results showed that the entorhinal area ranked as the fourth discriminative ROI with a high frequency of 84.5%, confirming that the entorhinal area is highly related to TS.

The SWM beneath the inferior frontal gyrus (IFWM) is the third discriminative ROI and TS group showed significantly reduced FA compared to control group. Concordantly, the previous study [48] also found that patients with TS showed significantly reduced FA in the left inferior frontal gyrus (below BA 44). Apart from IFWM, the left MFWM beneath the middle frontal gyrus, which was also found to be related to TS [49], is another frontal SWM region ranking within the top 22 discriminative ROI.

In addition to the frontal SWM, we also found that FA values and GM volumes significantly changed between groups in the postcentral WM and postcentral gyrus. The postcentral WM is the WM below the postcentral (BA 3a) cortex. In line with our finding, Thomalla et al. [50] found that FA values changed between groups in the postcentral WM. Other regions of interest involved the superior temporal WM beneath the superior temporal gyrus, which were significantly correlated with

tic occurrence in TS patients [51], and the inferior occipital WM beneath the inferior occipital gyrus, which is involved in the pathophysiology of TS [52].

For gyral regions with significant change in GM volume, the supramarginal gyrus and the inferior temporal cortex also well-correlated with TS [53].

4.3.2. Deep gray matter

Some discriminative ROIs in the basal ganglia (globus pallidus, putamen) and the limbic system (amygdala) also ranked within the top 22. In primates, the basal ganglia include the striatum (caudate nucleus and putamen), subthalamic nucleus (STN), globus pallidus internal and external segments, and substantia nigra pars compacta and pars reticulata. The basal ganglia were constantly associated with TS pathophysiology in children using VBM [54,55], with the same variation trend of GMP as our result. In addition, our results showed significantly increased AD value in the amygdalae, suggesting that altered connectivity of this limbic structure highly involved in the processing of memory, decision-making, and emotional reactions, which may play an important role in the pathophysiology of TS [56].

4.3.3. Deep white matter

The left superior cerebellar peduncle (SCP) in deep WM is the second most discriminative ROI feature with a high frequency of 99.0%. The SCP is the primary output of the cerebellum with mostly fibers carrying information to the midbrain, which was found to be connected to areas of altered FA in TS patients [50]. The SCP represents the major output of the cerebellum, including cerebello-thalamo-cortical pathways involved in sensory-motor integration [57]. Our results also indicate that SCP is an important connectivity node of the somatosensory system [58]. The middle cerebellar peduncle (MCP) ranked as the fifth discriminative regions; it is also an important output of the cerebellum, which connects the cerebellum to the pons. Consequently, the significant microstructure changes in SCP and MCP might suggest that the dysfunction of neural circuit involving cerebellum played a unique role in the TS pathology.

The superior corona radiata is one of the projection fibers, which are afferents carrying information to the cerebral cortex. Remarkably, the AD increase started from the corona radiata and extended to the anterior thalamic radiation [24], thereby indicating that alterations in

Tourette syndrome reach beyond the motor system [59]. Superior longitudinal fasciculus is long association fiber bundles. The modulation of tics by cortical activation depends on the long association fiber bundles within one hemisphere [60]. Our results conform to the previous study with TBSS which showed significant AD increased in these WM tracts in TS children [24].

Therefore, the consistency of our results with the previous findings suggests the effectiveness of our method in identifying potential biomarkers for TS classification.

5. Limitations

Our study has several limitations. First, the number of subjects in our study is relatively small. Second, our study includes several TS patients with ADHD, which may bias our results to some extent. Future studies with larger sample size and TS patients without ADHD are warranted to establish a more reliable and robust classification framework for both TS diagnosis and discriminative brain regions identification. Another limitation is that our paper focuses on using traditional machine-learning methods to establish a robust TS diagnostic model, several other machine-learning methods (i.e. CCA and co-regularization) will be explored in our future work. In particular, deep learning methods have recently made notable advances in the tasks of classification, regression and prediction in the neuroimaging research field. Although we did not present the results based on deep learning methods in this paper, it will be the focus of our future work to explore high-order deep learned features, to improve the accuracy of TS classification.

6. Conclusions

Overall, our study demonstrates the efficacy of using multi-modal MRI in conjunction with SVM and MKL classifiers for TS subject-specific diagnosis. At present, we achieved higher accuracy for TS classification than previous studies. Besides, the fully automated procedure of the proposed classification framework can be of great assistance in clinical TS diagnosis. Our framework identified key discriminative regions that are altered in TS patients, and all the spotted regions were in line with the state-of-the-art. Therefore, our study may provide some potential biomarkers for TS diagnosis and help reveal the pathological mechanism of TS. The proposed framework may be useful for early identification of TS subjects, and holds promise for predicting prognosis and treatment outcome in TS individuals.

Acknowledgments

We thank Dr. Hao Huang at University of Pennsylvania for consultation and support on MR pulse sequences.

This work was supported by National Natural Science Foundation of China (61271151, 91520202, 31271161), Youth Innovation Promotion Association CAS and Beijing Municipal Administration of Hospitals Incubating Program (PX2016035), Beijing Health System Top Level Health Technical Personnel Training Plan (2015-3-082).

References

- [1] R.J. Felling, H.S. Singer, Neurobiology of Tourette syndrome: current status and need for further investigation, *J. Neurosci.* 31 (2011) 12387–12395.
- [2] A. Stokes, H.N. Bawden, P.R. Camfield, J.E. Backman, J.M. Dooley, Peer problems in Tourettes disorder, *Pediatrics* 87 (1991) 936–942.
- [3] E. Gerard, B.S. Peterson, Developmental processes and brain imaging studies in Tourette syndrome, *J. Psychosom. Res.* 55 (2003) 13–22.
- [4] H.S. Singer, K. Minzer, Neurobiology of Tourette's syndrome: concepts of neuroanatomic localization and neurochemical abnormalities, *Brain Dev.* 25 (2003) S70–S84.
- [5] M.I. Makki, M. Behen, A. Bhatt, B. Wilson, H.T. Chugani, Microstructural abnormalities of striatum and thalamus in children with Tourette syndrome, *Mov. Disord.: Off. J. Mov. Disord. Soc.* 23 (2008) 2349–2356.
- [6] K.S. McNaught, J.W. Mink, Advances in understanding and treatment of Tourette syndrome, *Nat. Rev. Neurol.* 7 (2011) 667–676.
- [7] C. Ganos, V. Roessner, A. Munchau, The functional anatomy of Gilles de la Tourette syndrome, *Neurosci. Biobehav. Rev.* 37 (2013) 1050–1062.
- [8] L. O'Dwyer, F. Lamberton, A.L.W. Bokde, M. Ewers, Y.O. Faluyi, C. Tanner, B. Mazoyer, D. O'Neill, M. Bartley, D.R. Collins, T. Coughlan, D. Prvulovic, H. Hampel, Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment, *Plos One* 7 (2012).
- [9] L. O'Dwyer, F. Lamberton, S. Matura, M. Scheibe, J. Miller, D. Rujescu, D. Prvulovic, H. Hampel, White matter differences between healthy young ApoE4 carriers and non-carriers identified with tractography and support vector machines, *Plos One* 7 (2012).
- [10] S. Calderoni, A. Retico, L. Biagi, R. Tancredi, F. Muratori, M. Tosetti, Female children with autism spectrum disorder: an insight from mass-univariate and pattern classification analyses, *Neuroimage* 59 (2012) 1013–1022.
- [11] F. Liu, C.Y. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification, *Neuroimage* 84 (2014) 466–475.
- [12] Z. Dai, C. Yan, Z. Wang, J. Wang, M. Xia, K. Li, Y. He, Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3), *Neuroimage* 59 (2012) 2187–2195.
- [13] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, C. Plant, A. Oswald, T. Meindl, M. Pievani, A.L.W. Bokde, A. Fellgiebel, Combining DTI and MRI for the Automated Detection of Alzheimer's Disease Using a Large European Multicenter Dataset, Springer: Berlin Heidelberg, 2012.
- [14] D. Zhang, D. Shen, I. Alzheimer's Disease Neuroimaging, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, *Neuroimage* 59 (2012) 895–907.
- [15] D. Dai, J. Wang, J. Hua, H. He, Classification of ADHD children through multimodal magnetic resonance imaging, *Front. Syst. Neurosci.* 6 (2012) 63.
- [16] W. Liao, Y. Yu, H.-H. Miao, Y.-X. Feng, G.-J. Ji, J.-H. Feng, Inter-hemispheric intrinsic connectivity as a neuromarker for the diagnosis of boys with Tourette syndrome, *Mol. Neurobiol.* (2016) 1–9.
- [17] P.J. Basser, Inferring microstructural features and the physiological state of tissues from diffusion-weighted images, *NMR Biomed.* 8 (1995) 333–344.
- [18] P.J. Basser, D.K. Jones, Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review, *NMR Biomed.* 15 (2002) 456–467.
- [19] S.K. Song, J. Yoshino, T.Q. Le, S.J. Lin, S.W. Sun, A.H. Cross, R.C. Armstrong, Demyelination increases radial diffusivity in corpus callosum of mouse brain, *Neuroimage* 26 (2005) 132–140.
- [20] J.M. Tyszka, C. Readhead, E.L. Bearer, R.G. Pautler, R.E. Jacobs, Statistical diffusion tensor histology reveals regional dysmyelination effects in the shiverer mouse mutant, *Neuroimage* 29 (2006) 1058–1065.
- [21] S.K. Song, S.W. Sun, W.K. Ju, S.J. Lin, A.H. Cross, A.H. Neufeld, Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia, *Neuroimage* 20 (2003) 1714–1722.
- [22] J. Ashburner, K.J. Friston, Voxel-based morphometry - the methods, *Neuroimage* 11 (2000) 805–821.
- [23] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C.E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, T.E. Behrens, Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data, *Neuroimage* 31 (2006) 1487–1505.
- [24] Y. Liu, W. Miao, J. Wang, P. Gao, G. Yin, L. Zhang, C. Lv, Z. Ji, T. Yu, B.A. Sabel, H. He, Y. Peng, Structural abnormalities in early Tourette syndrome children: a combined voxel-based morphometry and tract-based spatial statistics study, *Plos One* 8 (2013) e76105.
- [25] T.M. Nir, N. Jahanshad, J.E. Villalon-Reina, A.W. Toga, C.R. Jack, M.W. Weiner, P.M. Thompson, I. Alzheimer's Disease Neuroimaging, Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging, *Neuroimage Clin.* 3 (2013) 180–195.
- [26] S. Sonnenburg, G. Ratsch, C. Schafer, B. Scholkopf, Large scale multiple kernel learning, *J. Mach. Learn. Res.* 7 (2006) 1531–1565.
- [27] L. Seahill, M.A. Riddle, M. McSwiggin-Hardin, S.I. Ort, R.A. King, W.K. Goodman, D. Cicchetti, J.F. Leckman, Children's yale-brown obsessive compulsive scale: reliability and validity, *J. Am. Acad. Child Adolesc. Psychiatry* 36 (1997) 844–852.
- [28] P. Retz-Junginger, W. Retz, D. Blocher, R.D. Stieglitz, T. Georg, T. Supprian, P.H. Wender, M. Rosler, Reliability and validity of the Wender-Utah-Rating-Scale short form. Retrospective assessment of symptoms for attention deficit/hyperactivity disorder, *Der Nervenarzt* 74 (2003) 987–993.
- [29] J.F. Leckman, M.A. Riddle, M.T. Hardin, S.I. Ort, K.L. Swartz, J. Stevenson, D.J. Cohen, The yale global tic severity scale: initial testing of a clinician-rated scale of tic severity, *J. Am. Acad. Child Adolesc. Psychiatry* 28 (1989) 566–573.
- [30] J. Ashburner, A fast diffeomorphic image registration algorithm, *Neuroimage* 38 (2007) 95–113.
- [31] C.J. Holmes, R. Hoge, L. Collins, R. Woods, A.W. Toga, A.C. Evans, Enhancement of MR images using registration for signal averaging, *J. Comput. Assist. Tomogr.* 22 (1998) 324–333.
- [32] J.D. Schmahmann, J. Doyon, D. McDonald, C. Holmes, K. Lavoie, A.S. Hurwitz, N. Kabani, A. Toga, A. Evans, M. Petrides, Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space, *Neuroimage* 10 (1999) 233–260.
- [33] K. Shiva, N.S. Ryan, I.B. Malone, M. Marc, C. David, G.R. Ridgway, Z. Hui, N.C. Fox, O. Sebastian, The importance of group-wise registration in tract based spatial statistics study of neurodegeneration: a simulation study in Alzheimer's disease, *Plos One* 7 (2012) e45996.

- [34] K. Oishi, A. Faria, H.Y. Jiang, X. Li, K. Akhter, J.Y. Zhang, J.T. Hsu, M.I. Miller, P.C.M. van Zijl, M. Albert, C.G. Lyketsos, R. Woods, A.W. Toga, G.B. Pike, P. Rosa-Neto, A. Evans, J. Mazziotta, S. Mori, Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants, *Neuroimage* 46 (2009) 486–499.
- [35] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1997) 273–324.
- [36] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, *Mach. Learn.* 46 (2002) 389–422.
- [37] S.M. Wilson, J.M. Ogar, V. Laluz, M. Growdon, J. Jang, S. Glenn, B.L. Miller, M.W. Weiner, M.L. Gorno-Tempini, Automated MRI-based classification of primary progressive aphasia variants, *Neuroimage* 47 (2009) 1558–1567.
- [38] C. Hinrichs, V. Singh, G.F. Xu, S.C. Johnson, A.D. Neuroimaging, Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population, *Neuroimage* 55 (2011) 574–589.
- [39] A. Rakotomamonjy, F.R. Bach, S. Canu, Y. Grandvalet, SimpleMKL, *J. Mach. Learn. Res.* 9 (2008) 2491–2521.
- [40] S.V.N. Vishwanathan, Z. Sun, N. Ampornpant, M. Varma, Multiple Kernel Learning and the SMO Algorithm, *Jmlr*, 7 (2010) 2006.
- [41] H. Wen, Y. Liu, J. Wang, I. Rekik, J. Zhang, Y. Zhang, H. Tian, Y. Peng, H. He, Combining tract- and atlas-based analysis reveals microstructural abnormalities in early Tourette syndrome children, *Human brain mapping*, 2016.
- [42] M. Ingalhalikar, S. Kanterakis, R. Gur, T.P. Roberts, R. Verma, DTI based diagnostic prediction of a disease via pattern classification, Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 13, 2010, pp. 558–565.
- [43] A. Fornito, A. Zalesky, C. Pantelis, E.T. Bullmore, Schizophrenia, neuroimaging and connectomics, *Neuroimage* 62 (2012) 2296–2314.
- [44] A. Nazeri, M.M. Chakravarty, D. Felsky, N.J. Lobaugh, T.K. Rajji, B.H. Mulsant, A.N. Voineskos, Alterations of superficial white matter in schizophrenia and relationship to cognitive performance, *Neuropsychopharmacol* 38 (2013) 1954–1962.
- [45] G. Leisman, C. Machado, R. Melillo, R. Mualem, Intentionality and 'free-will' from a neurodevelopmental perspective, *Front. Integr. Neurosci.* 6 (2012) 36.
- [46] G. Gravano, Gilles de la Tourette syndrome, *Ann. Clin. Psychiatry: Off. J. Am. Acad. Clin. Psychiatr.* 25 (2013) 297–306.
- [47] Z.S. Wang, T.V. Maia, R. Marsh, T. Colibazzi, A. Gerber, B.S. Peterson, The neural circuits that generate tics in Tourette's syndrome, *Am. J. Psychiatry* 168 (2011) 1326–1337.
- [48] K.R. Muller-Vahl, J. Grosskreutz, T. Prell, J. Kaufmann, N. Bodammer, T. Peschel, Tics are caused by alterations in prefrontal areas, thalamus and putamen, while changes in the cingulate gyrus reflect secondary compensatory mechanisms, *BMC Neurosci.* 15 (2014) 6.
- [49] K.R. Muller-Vahl, J. Kaufmann, J. Grosskreutz, R. Dengler, H.M. Emrich, T. Peschel, Prefrontal and anterior cingulate cortex abnormalities in Tourette syndrome: evidence from voxel-based morphometry and magnetization transfer imaging, *BMC Neurosci.* 10 (2009) 47.
- [50] G. Thomalla, H.R. Siebner, M. Jonas, T. Baumer, K. Biermann-Ruben, F. Hummel, C. Gerloff, K. Muller-Vahl, A. Schnitzler, M. Orth, A. Munchau, Structural changes in the somatosensory system correlate with tic severity in Gilles de la Tourette syndrome, *Brain* 132 (2009) 765–777.
- [51] E. Stern, D.A. Silbersweig, K.Y. Chee, A. Holmes, M.M. Robertson, M. Trimble, C.D. Frith, R.S.J. Frackowiak, R.J. Dolan, A functional neuroanatomy of tics in Tourette syndrome, *Arch. Gen. Psychiatry* 57 (2000) 741–748.
- [52] B.S. Peterson, L. Staib, L. Scabill, H. Zhang, C. Anderson, J.F. Leckman, D.J. Cohen, J.C. Gore, J. Albert, R. Webster, Regional brain and ventricular volumes in Tourette syndrome, *Arch. Gen. Psychiatry* 58 (2001) 427–440.
- [53] K.J. Jeffries, C. Schooler, C. Schoenbach, P. Herscovitch, T.N. Chase, A.R. Braun, The functional neuroanatomy of Tourette's syndrome: an FDG PET study III: functional coupling of regional cerebral metabolic rates, *Neuropsychopharmacol* 27 (2002) 92–104.
- [54] V. Roessner, S. Overlack, C. Schmidt-Samoa, J. Baudewig, P. Dechent, A. Rothenberger, G. Helms, Increased putamen and callosal motor subregion in treatment-naïve boys with Tourette syndrome indicates changes in the bihemispheric motor network, *J. Child Psychol. Psychiatry Allied Discip.* 52 (2011) 306–314.
- [55] A.G. Ludolph, F.D. Juengling, G. Libal, A.C. Ludolph, J.M. Fegert, J. Kassubek, Grey-matter abnormalities in boys with Tourette syndrome: magnetic resonance imaging study using optimised voxel-based morphometry, *Br. J. Psychiatry: J. Ment. Sci.* 188 (2006) 484–485.
- [56] C.J. Werner, T. Stocker, T. Kellermann, H.P. Wegener, F. Schneider, N.J. Shah, I. Neuner, Altered amygdala functional connectivity in adult Tourette's syndrome, *Eur. Arch. Psychiatry Clin. Neurosci.* 260 (Suppl. 2) (2010) S95–S99.
- [57] L.N. Hazrati, A. Parent, Projection from the deep cerebellar nuclei to the pedunculopontine nucleus in the squirrel monkey, *Brain Res.* 585 (1992) 267–271.
- [58] G. Thomalla, H.R. Siebner, M. Jonas, T. Baumer, K. Biermann-Ruben, F. Hummel, C. Gerloff, K. Muller-Vahl, A. Schnitzler, M. Orth, A. Munchau, Structural changes in the somatosensory system correlate with tic severity in Gilles de la Tourette syndrome, *Brain* 132 (2009) 765–777.
- [59] I. Neuner, Y. Kupriyanova, T. Stocker, R. Huang, O. Posnansky, F. Schneider, M. Tittgemeyer, N.J. Shah, White-matter abnormalities in Tourette syndrome extend beyond motor pathways, *Neuroimage* 51 (2010) 1184–1193.
- [60] J.A. Church, D.A. Fair, N.U.F. Dosenbach, A.L. Cohen, F.M. Miezin, S.E. Petersen, B.L. Schlaggar, Control networks in paediatric Tourette syndrome show immature and anomalous patterns of functional connectivity, *Brain* 132 (2009) 225–238.

Hongwei Wen received B.Eng. degrees in Software Engineering from Wu Yuzhang Honors College of Sichuan University (2011). He is now working towards the Ph.D. degree in Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences. His research interests include MRI, medical image analysis, machine learning and pattern recognition.

Yue Liu, MD, Ph.D., associate Professor, Radiologist in Department of Radiology, Beijing Children's Hospital, Capital Medical University, Beijing, China. She has become an academic staff in radiology department since 1992. Her research interests are focusing on pediatric imaging and fMRI study.

Islem Rekik was awarded a Ph.D. in "Neuroimaging and Computer Sciences" from the University of Edinburgh (UK) in 2014. Her current research work focuses on understanding postnatal brain development as well as modeling and predicting brain diseases and disorders evolution using advanced machine learning and computational anatomy analysis methods.

Shengpei Wang received B. Eng. degree from East China University of Science and Technology (2013). He is now working towards the Ph.D. degree in Institute of Automation, Chinese Academy of Sciences. His research interests include medical image analysis and Brain-computer interface (BCI).

Zhiqiang Chen received B. Eng. degrees from University of Science and Technology, Beijing (2014). He is now working towards the master degree in Institute of Automation, Chinese Academy of Sciences. His research interests include Deep Learning, Computer Vision, medical image analysis, machine learning and pattern recognition.

Jishui Zhang is an associate chief physician at Beijing Children's Hospital, Capital Medical University, Beijing, China. He received his Ph.D. in 2004 from Xiangya medical school of Medicine, Central South University. His research interests include the study of neurology and psychology of diseases.

Yue Zhang is a technician of MRI at Beijing Children's Hospital, Capital Medical University, Beijing, China. She received university degree from Jilin University.

Yun Peng, MD, Ph.D., Professor, Chair of Department of Radiology, Beijing Children's Hospital, Capital Medical University, Beijing, China. She has become an academic staff in the above department since 1995. Prof. Peng's researches are focusing on pediatric imaging.

Huiguang He received BS degree (1994) and MS degree (1997) from Dalian Maritime University, Liaoning, China. He was awarded his Ph.D. degree with honor from Institute of Automation, Chinese Academy of Sciences (CASIA) in 2002. Then he joined CASIA as faculty member and is currently a full professor. His research interest includes pattern recognition, medical image processing and brain computer interface.