



# Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification



Jin Liu<sup>a</sup>, Yi Pan<sup>b</sup>, Fang-Xiang Wu<sup>c</sup>, Jianxin Wang<sup>a,\*</sup>

<sup>a</sup>Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China

<sup>b</sup>Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA

<sup>c</sup>Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon S7N5A9, Canada

## ARTICLE INFO

### Article history:

Received 18 January 2019

Revised 27 February 2020

Accepted 8 March 2020

Available online 14 March 2020

Communicated by Dr. Nianyin Zeng

### Keywords:

Multi-modal MRI data

Multi-view information

Multi-task feature selection

Multi-kernel learning

MCI classification

## ABSTRACT

The classification of mild cognitive impairment (MCI), which is a early stage of Alzheimer's disease and is associated with brain structural and functional changes, is still a challenging task. Recent studies have shown great promise for improving the performance of MCI classification by combining multiple structural and functional features, such as grey matter volume and clustering coefficient. However, extracting which features and how to combine multiple features to improve the performance of MCI classification have always been challenging problems. To address these problems, in this study we propose a new method to enhance the feature representation of multi-modal MRI data by combining multi-view information to improve the performance of MCI classification. Firstly, we extract two structural features (including grey matter volume and cortical thickness) and two functional features (including clustering coefficient and shortest path length) of each cortical brain region based on automated anatomical labeling (AAL) atlas from both T1w MRI and rs-fMRI data of each subject. Then, in order to obtain features that are more helpful in distinguishing MCI subjects, an improved multi-task feature selection method, namely MTFs-gLASSO-TTR, is proposed. Finally, a multi-kernel learning algorithm is adopted to combine multiple features to perform the MCI classification task. Our proposed MCI classification method is evaluated on 315 subjects (including 105 LMCI subjects, 105 EMCI subjects and 105 NCs) with both T1w MRI and rs-fMRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results show that our proposed method achieves an accuracy of 88.5% and an area under the receiver operating characteristic (ROC) curve (AUC) of 0.897 for LMCI/NC classification, an accuracy of 82.7% and an AUC of 0.832 for EMCI/NC classification, and an accuracy of 79.6% and an AUC of 0.803 for LMCI/EMCI classification, respectively. In addition, by comparison, the accuracy and AUC values of our proposed method are better than those of some existing state-of-the-art methods in MCI classification. Overall, our proposed MCI classification method is effective and promising for automatic diagnosis of MCI in clinical practice.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

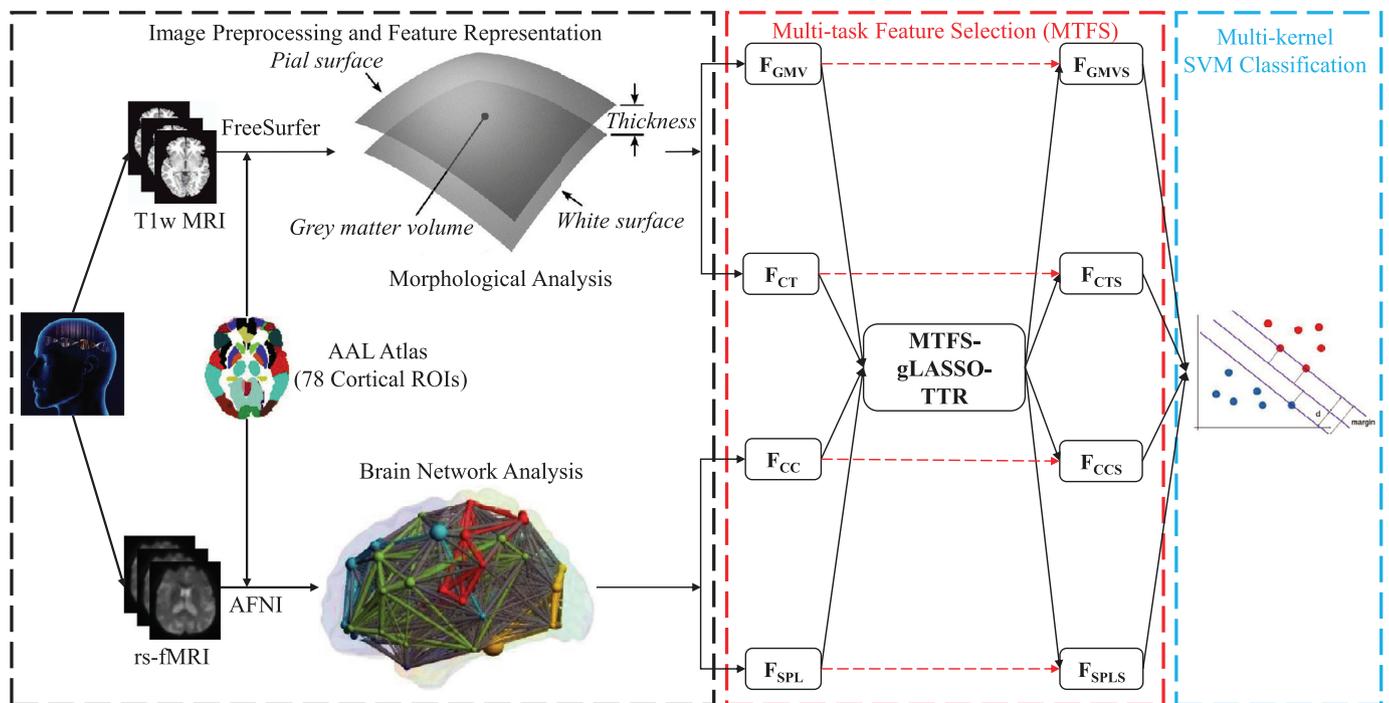
Alzheimer's disease (AD) is a common neurodegenerative disorder. Its clinical manifestations are decreased memory, persistent cognitive decline, motor impairment and so on. As of 2006, there are 26.6 million AD patients worldwide, and by 2050, one in every 85 people worldwide is expected to have AD [1]. As the world is developing into an aging society, the negative impact of AD on families and society will become more and more significant. Mild cognitive impairment (MCI) is an intermediate process in the conversion of normal people to AD, with up to 15% of people with MCI

being converted to AD each year [2–4]. At present, there is no accurate diagnosis and effective treatment for AD. Most researchers hope that patients can be discovered in time when they are in the stage of MCI, and then take effective measures to prevent further deterioration of the disease. Therefore, accurate early classification of MCI subjects is very important for human health.

Since magnetic resonance imaging (MRI) can noninvasively measure brain structural and functional changes related to brain disorder development *in vivo*, in recent years it has been widely used in the study of brain disorders, such as AD/MCI [5,6], schizophrenia [7,8], autism [9,10] and so on. Therefore, MRI can provide phenotypes that can be used to diagnose such disorders. MRI falls into two broad categories: structural MRI (such as T1 MRI, T2 MRI and so on.) and functional MRI (such as rs-fMRI, ts-fMRI and so on.). Brain structure is typically measured using

\* Corresponding author.

E-mail address: [jxwang@mail.csu.edu.cn](mailto:jxwang@mail.csu.edu.cn) (J. Wang).



**Fig. 1.** An overall flowchart for MCI classification by combining multi-view information from multi-modal MRI data, which consists of three main steps: (1) image preprocessing and feature representation, (2) multi-task feature selection (MTFS), and (3) multi-kernel SVM classification.

structural MRI, which can provide relatively high-definition brain structure in grey matter and white matter. There are many metrics to measure brain structure, and most of them have been widely applied in the study of AD and MCI, such as grey matter volume, cortical thickness, texture properties and so on [11–15]. Brain function is typically measured using functional MRI, which can provide changes in hemodynamics caused by neuronal activity. Functional connectivity between brain regions is a common measure of brain function. Also, brain networks based on brain regions and functional connectivity between brain regions have been widely used for feature representation in the study of various brain disorders. In the past years, brain function analysis based on graph theory has shown a powerful role in exploring functional impairment of brain disorders, and has been widely used in the study of AD and MCI [16–19].

In the past decade, whether structural MRI-based brain structure metrics or functional MRI-based brain function metrics, these metrics were mainly used separately in the studies with MCI. For example, Karas et al. [11] found that the MCI subjects showed a decrease in grey matter volume in the medial temporal lobe. Wang et al. [18] constructed functional brain networks of MCI subjects and found that the length of the shortest path increased in MCI subjects compared with NCs; Zhang et al. [20] first extracted functional connectivity between brain regions from functional MRI data of each subject as feature representation, and then trained a L2-regularized logistic regression classifier based on these functional connectivity features to perform MCI classification. Therefore, many researchers believe that different metrics may contain different-yet-complementary information, and combinations of these metrics may improve MCI classification performance over separate metrics. In fact, recent studies have also shown great promise for improving the accuracy of MCI classification by combining multiple structural and functional metrics, such as grey matter volume, clustering coefficient, cortical thickness and so on. For example, Wee et al. [21] first used both structural MRI and functional MRI data of each subject to construct multiple brain

networks for each subject, and then extracted local clustering coefficient from each brain network of each subject as feature representation to perform the MCI classification task by using a multi-kernel learning algorithm; De Marco et al. [22] used multiple machine learning models based on different metrics from both structural MRI and functional MRI data to investigate the performance of MCI classification; Tripathi et al. [23] proposed an unsupervised framework for the classification of EMCI and LMCI by combining shape and voxel-based features from 12 brain regions; Jie et al. [24] proposed a feature combination framework to combine both temporal and spatial features of dynamic functional networks to perform automatic classification of EMCI and LMCI. Recently, with the development of deep learning [25–28], especially convolutional neural networks [29–31], some researchers have begun to use convolutional neural networks to directly extract features from structural MRI and functional MRI images for AD or MCI classification [32–34]. For example, Islam and Zhang [32] proposed a deep convolutional neural network for AD classification using brain structural MRI images; Sarraf and Tofghi [33] employed a convolutional neural network to distinguish an AD brain from a healthy brain using functional MRI images. So far, although some results have been achieved for the classification of MCI subjects based on structural and functional MRI data, extracting which features and how to combine multiple features to improve MCI classification accuracy have always been difficult problems.

In order to try to address these two problems, in this study we propose a new method for MCI classification by combining multi-view information from multi-modal MRI data, which is shown in Fig. 1. Firstly, we measure two regional structural features including grey matter volume and cortical thickness based on automated anatomical labeling (AAL) atlas [35] from T1w MRI data for each subject. Then, we construct an individual brain network based on AAL atlas for rs-fMRI data of each subject, and adopt brain network analysis method to compute two regional functional features including clustering coefficient and the shortest path length for each subject. Therefore, we can obtain four feature sets including two

**Table 1**  
Demographic information of the subjects in this study from ADNI database.

Group	Number	Age	Gender (M/F)	MMSE
NC	105	77.1 ± 6.3	54/51	29.1 ± 1.1
EMCI	105	76.3 ± 5.4	49/56	27.5 ± 1.8
LMCI	105	75.8 ± 6.3	35/70	26.6 ± 2.2

regional structural feature sets from T1w MRI data and two regional functional feature sets from rs-fMRI data for each subject. Next, in order to obtain features that are as effective as possible for MCI classification, an improved multi-task feature selection method, denoted as MTFs-gLASSO-TTR, is proposed and applied to the above-mentioned four feature sets. Finally, in order to perform the MCI classification task, a multi-kernel learning algorithm is applied to combine the selected four feature sets. Our proposed MCI classification method is evaluated on 315 subjects (including 105 LMCI subjects, 105 EMCI subjects and 105 NCs) with T1w MRI and rs-fMRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>).

## 2. Materials and method

### 2.1. Subjects

A subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) [36] is used to evaluate our proposed MCI classification method. This subset includes 315 subjects with both T1w MRI and rs-fMRI data, which are composed of 105 subjects with late mild cognitive impairment (LMCI), 105 subjects with early mild cognitive impairment (EMCI), and 105 normal controls (NCs). All T1w MRI and rs-fMRI data are acquired on 3.0 Tesla Philips medical system scanners at multiple sites, and the slice thickness of T1w MRI data and rs-fMRI data is 1.2 mm and 3.0 mm, respectively. Furthermore, rs-fMRI data of each subject contain 140 volumes. The demographic information of the studied subjects is presented in Table 1, where MMSE is the abbreviation of Mini Mental State Examination, and the front and back of ± represent mean and standard deviation, respectively. For more details with these subjects from ADNI, please see <http://adni.loni.usc.edu/>.

### 2.2. Method

#### 2.2.1. Image preprocessing and feature representation

Fig. 1 shows the procedures of image preprocessing and feature representation for both T1w MRI and rs-fMRI data. As can be seen from Fig. 1, the procedures of image preprocessing and feature representation mainly include two aspects: T1w MRI data preprocessing and feature representation, and rs-fMRI data preprocessing and feature representation. These two aspects are briefly introduced as follows.

Firstly, a standard preprocessing procedure is applied to T1 MRI data of each subject using a standard FreeSurfer pipeline (<https://surfer.nmr.mgh.harvard.edu>) [37], including motion correction, non-uniform intensity normalization, talairach transform computation, skull removal, volumetric segmentation, cortical surface reconstruction and so on. After this standard preprocessing procedure, we can obtain two boundaries: the white surface boundary (i.e., grey matter-white matter interface) and the pial surface boundary (i.e., grey matter-cerebrospinal fluid interface) from T1w MRI data of each subject, which are shown in Fig. 1. Based on these two boundaries of grey matter, we can measure morphological information on the grey matter of each subject, such as grey matter volume (GMV), cortical thickness (CT) and so on. Previous studies have shown that GMV and CT have been widely applied in the morphological analysis of MCI and AD [38,39]. For those

reasons, in this study we extract these two regional morphological measures based on each cortical region of the AAL atlas as structural feature representation from T1w MRI data for each subject. All cortical regions of the AAL atlas are shown in Table 2. Therefore, we can obtain two regional morphological feature sets from T1w MRI data for each subject, which are denoted as  $F_{GMV}$  and  $F_{CT}$ , respectively. It is worth mentioning that these two regional morphological feature sets are all 78-dimensional vectors.

Secondly, as can be seen from Fig. 1, a standard preprocessing procedure is also applied to rs-fMRI data of each subject using the pipeline provided by the Analysis of Functional Neuroimages (AFNI) software (<https://afni.nimh.nih.gov/>) [40], including removing the first 10 rs-fMRI volumes, slice timing, head motion corrections, spatial smoothing, band-pass filtering (0.01–0.1 Hz), nuisance signal regression, and Montreal Neurological Institute (MNI) space normalization and so on. After this standard preprocessing procedure, in this study we parcellate the resulted rs-fMRI data of each subject into 78 cortical regions (as shown in Table 2) according to the AAL atlas. Therefore, we can obtain the average rs-fMRI time series of each cortical region for each subject. In order to consider functional connectivity between each two cortical regions for each subject, in this study the pairwise Pearson correlation coefficients between cortical regions  $i$  and  $j$  are defined and calculated by the following formula, i.e.,

$$F_c(i, j) = \frac{\text{cov}(s_i, s_j)}{\sigma_{s_i} \sigma_{s_j}} \quad (1)$$

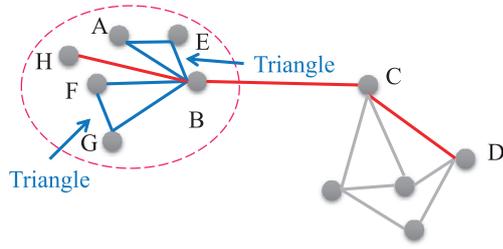
where  $F_c(i, j)$  denotes the connectivity weight between two cortical regions  $i$  and  $j$ ,  $\text{cov}(\cdot, \cdot)$  denotes the covariance between two vectors,  $s_i$  and  $s_j$  denote the average rs-fMRI time series of cortical regions  $i$  and  $j$ , respectively, and  $\sigma_{s_i}$  and  $\sigma_{s_j}$  denote the standard deviation of vectors  $s_i$  and  $s_j$ , respectively. So far, we can construct an individual brain network, which consists of 78 cortical regions according to the AAL atlas and functional connectivity between each two cortical regions, to represent rs-fMRI data of each subject.

At present, brain network analysis based on graph theory [41,42] plays an important role in the study of complex brain diseases, and is widely used in various brain diseases [43], such as AD/MCI, schizophrenia, Parkinson and so on. Many brain diseases have been found to be associated with the abnormal topological properties (such as clustering coefficient (CC) and shortest path length (SPL)) of brain networks, which have been widely used to brain disease diagnosis. For example, Yao et al. [44] found greater CC in MCI group compared with NC group, and Wang et al. [18] found increased SPL in MCI group compared with NC group. The small-world networks first proposed by Watts and Strogatz [45] have both high CC similar to regular networks and shorter SPL similar to random networks. In other words, The small-world networks combine the respective topological advantages of both regular networks and random networks to ensure the efficiency of information transmission at both local and global levels. Many researchers believe that brain diseases may destroy the small-world nature of brain networks, i.e. abnormal CC or SPL, or even both. For that reason, in this study we compute two regional network properties including CC and SPL based on individual brain network of each subject from rs-fMRI data, which are regarded as functional feature representation. These two regional network properties are briefly introduced as follows.

- The regional CC proposed by Onnela et al. [46] quantifies the presence of clusters that reflects functional segregation among nodes. A larger regional CC implies more functional segregation that allows the corresponding node and its neighbors for specialized information processing. When calculating the regional CC, in this study we only consider triplets of nodes. The regional CC based on a weighted undirected network can be

**Table 2**  
Cortical regions in the AAL atlas.

No. (L/R)	Regions	Abbr.	No. (L/R)	Regions	Abbr.
1/2	Precentral gyrus	PrCG	3/4	Superior frontal gyrus	SFG
5/6	Superior frontal gyrus, orbital part	SFGorb	7/8	Middle frontal gyrus	MiFG
9/10	Middle frontal gyrus, orbital part	MFGorb	11/12	Inferior frontal gyrus, pars opercularis	IFGope
13/14	Inferior frontal gyrus, pars triangularis	IFGtri	15/16	Inferior frontal gyrus, pars orbitalis	IFGorb
17/18	Rolandic operculum	ROL	19/20	Supplementary motor area	SMA
21/22	Olfactory cortex	OLF	23/24	Medial frontal gyrus	MeFG
25/26	Medial orbitofrontal cortex	MOC	27/28	gyrus rectus	REC
29/30	Insula	INS	31/32	Anterior cingulate gyrus	ACG
33/34	Middle cingulate gyrus	MCG	35/36	Posterior cingulate gyrus	PCG
37/38	Parahippocampal gyrus	PHG	39/40	Calcarine sulcus	CAL
41/42	Cuneus	CUN	43/44	Lingual gyrus	LING
45/46	Superior occipital gyrus	SOG	47/48	Middle occipital gyrus	MOG
49/50	Inferior occipital gyrus	IOG	51/52	Fusiform gyrus	FFG
53/54	Postcentral gyrus	PoCG	55/56	Superior parietal lobule	SPG
57/58	Inferior parietal lobule	IPL	59/60	Supramarginal gyrus	SMG
61/62	Angular gyrus	ANG	63/64	Precuneus	PCUN
65/66	Paracentral lobule	PCL	67/68	Transverse temporal gyrus	TTG
69/70	Superior temporal gyrus	STG	71/72	Superior temporal pole	STP
73/74	Middle temporal gyrus	MTG	75/76	Middle temporal pole	MTP
77/78	Inferior temporal gyrus	ITG			



**Fig. 2.** Illustration of CC and SPL in a simple weighted undirected network.

defined as

$$CC^w(i) = \frac{2}{k_i(k_i - 1)} \sum_{j,h} (\bar{w}_{ij}\bar{w}_{ih}\bar{w}_{jh})^{1/3} \quad (2)$$

where  $w_{ij}$ ,  $w_{ih}$ , and  $w_{jh}$  are the connection weights between nodes  $i$  and  $j$ , between nodes  $i$  and  $h$ , and between nodes  $j$  and  $h$ , respectively. The weights  $\bar{w}_{ij}$ ,  $\bar{w}_{ih}$  and  $\bar{w}_{jh}$  have been scaled, namely  $\bar{w}_{ij} \leftarrow w_{ij}/\max(w)$ ,  $\bar{w}_{ih} \leftarrow w_{ih}/\max(w)$ ,  $\bar{w}_{jh} \leftarrow w_{jh}/\max(w)$ ,  $\max(w)$  denotes the maximum connectivity weight in the brain network. The number of edges connected to node  $i$  is represented by  $k_i$ . For example, in Fig. 2, the CC of node B is calculated by  $CC^w(B) = \frac{2}{6(6-1)} \left( \frac{(w_{BA}w_{BE}w_{AE})^{1/3}}{\max\{w_{BA}, w_{BE}, w_{AE}\}} + \frac{(w_{BF}w_{BG}w_{FG})^{1/3}}{\max\{w_{BF}, w_{BG}, w_{FG}\}} \right)$ .

The SPL plays an important role in the information transmission of a brain network, and it is a very important metric to describe the internal structure of the brain network. Meanwhile, the SPL is a measure of functional integration, which can measure the ability to rapidly combine pieces of specialized information. The shorter SPL can transmit the information more quickly and reduce brain consumption. In a weighted undirected network, a path between nodes  $i$  and  $j$  with the minimum weight of edges is called the shortest path between these two nodes, and its length  $SPL(i, j)$  is denoted as

$$SPL^w(i, j) = \sum_{w_{pq} \in SP_{i \rightarrow j}} w_{pq} \quad (3)$$

where  $SP_{i \rightarrow j}$  is the shortest path between nodes  $i$  and  $j$ . For example, in Fig. 2, the SPL between nodes H and D is calculated by  $SPL^w(H, D) = w_{HB} + w_{BC} + w_{CD}$ . The SPL of node  $i$  can be represented by the average SPL between node  $i$  and all other

nodes, and is denoted as  $SPL^w(i)$ :

$$SPL^w(i) = \frac{1}{N-1} \sum_{j \neq i} SP_{i \rightarrow j} \quad (4)$$

Therefore, we can obtain two regional network feature sets from rs-fMRI data for each subject, and denoted as  $F_{CC}$  and  $F_{SPL}$ , respectively. It is worth mentioning that these two regional network feature sets are also all 78-dimensional vectors.

From the above, we can obtain four regional feature sets from both T1w MRI and rs-fMRI data for each subject, including two structural feature sets:  $F_{GMV}$  and  $F_{CT}$ , and two functional feature sets:  $F_{CC}$  and  $F_{SPL}$ . These four regional feature sets are taken as the original features of each subject.

### 2.2.2. Multi-task feature selection (MTFS)

LASSO (Least Absolute Shrinkage and Selection Operator) [47] is a regression analysis method and has been widely applied in feature selection. In the process of feature selection, the LASSO method can be formulated as

$$\min_{\mathbf{w}} (\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1) \quad (5)$$

where  $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^i, \dots, \mathbf{x}^N]^T \in \mathfrak{R}^{N \times P}$  denotes all training subjects,  $\mathbf{x}^i$  denotes the  $i$ th training subject,  $N$  denotes the number of the training subjects,  $P$  denotes the number of features of each training subject,  $\mathbf{y} = [y^1, y^2, \dots, y^i, \dots, y^N]^T \in \mathfrak{R}^N$  denotes the labels of all training subjects,  $\mathbf{w} \in \mathfrak{R}^P$  is the discriminant vector whose value indicates the contribution of each feature,  $\|\mathbf{w}\|_1$  denotes the  $l_1$ -norm of  $\mathbf{w}$ , and  $\lambda > 0$  is a parameter to balance the loss function (i.e.,  $\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2$ ) and the regularization term (i.e.,  $\|\mathbf{w}\|_1$ ). The larger the  $\lambda$  value, the greater the penalty for the parameters in the model, resulting in higher model sparsity, that is, more parameters are trained to zero.

Suppose there are  $T$  feature selection tasks similar to Eq. (5), these  $T$  feature selection tasks can be performed by Eq. (6) as follows,

$$\min_{\mathbf{w}_t} (\|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2^2 + \lambda \|\mathbf{w}_t\|_1), t = 1, 2, \dots, T. \quad (6)$$

where  $\mathbf{X}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^i, \dots, \mathbf{x}_t^N]^T \in \mathfrak{R}^{N \times P}$  denotes all training subjects in the  $t$ th task,  $\mathbf{y}_t = [y_t^1, y_t^2, \dots, y_t^i, \dots, y_t^N]^T \in \mathfrak{R}^N$  denotes the labels of all training subjects in the  $t$ th task, and  $\mathbf{w}_t \in \mathfrak{R}^P$  is the discriminant vector whose value indicates the contribution of each feature in the  $t$ th task. However, although this strategy can solve  $T$

learning tasks separately, it ignores the relationship between these feature selection tasks. In general, the feature selection methods in which tasks are usually learned independently of each other are called single-task feature selection (STFS) methods.

To further consider the relationship between different feature selection tasks, the different feature selection tasks should be learned jointly, which is often called multi-task feature selection (MTFS). At present, the group LASSO-based MTFS method (denoted as MTFS-gLASSO) [48] is a common feature selection method, and has been widely used in various feature selection tasks. The MTFS-gLASSO method can be formulated as follows,

$$\min_{\mathbf{W}} \left( \sum_{t=1}^T \|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} \right) \quad (7)$$

where  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i, \dots, \mathbf{w}_T] \in \mathbb{R}^{P \times T}$  is a discriminant matrix,  $\|\mathbf{W}\|_{2,1}$  denotes the  $l_{2,1}$ -norm of  $\mathbf{W}$ . However, Eq. (7) only considers the joint selection of the same features across different tasks, but the relationship between tasks themselves is still not taken into account. To address this problem, Wang et al. [10] proposed a regularizer based on task-task relationship (TTR), which is formulated as follows:

$$R_{t-t}(\mathbf{W}) = \frac{1}{2} \sum_{i \neq j} g_{i,j} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \quad (8)$$

$$g_{i,j} = \exp \left( \frac{-2 \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_2^2}{\sigma^2} \right) \quad (9)$$

$$\sigma^2 = \sum_{i=1}^T \sum_{j=1}^T \frac{\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_2^2}{T^2} \quad (10)$$

where  $\bar{\mathbf{x}}_i$  is the average vector of all training subjects in the  $i$ th task, and  $\bar{\mathbf{x}}_j$  is similar to  $\bar{\mathbf{x}}_i$ . By incorporating the TTR regularization into Eq. (7), we can obtain the following MTFS model, which is denoted as MTFS-gLASSO-TTR:

$$\min_{\mathbf{W}} \left( \sum_{t=1}^T \|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2^2 + \lambda \|\mathbf{W}\|_{2,1} + \beta \sum_{i,j} g_{i,j} \|\mathbf{w}_i - \mathbf{w}_j\|_2^2 \right) \quad (11)$$

where  $\beta$  is also a regularization parameter, and  $\beta > 0$ . It is worth mentioning that  $\lambda$  is used to control the number of selected features from the original features, while  $\beta$  is used to get more discriminative features from the selected features.

In this study, the MTFS-gLASSO-TTR method is used to select the more discriminant features from the above four regional feature sets (i.e.,  $F_{GMV}$ ,  $F_{CT}$ ,  $F_{CC}$  and  $F_{SPL}$ ) for MCI classification as shown in Fig. 1. After this MTFS step, we can obtain four selected feature sets for each training subject, which are denoted as  $F_{GMVS}$ ,  $F_{CTS}$ ,  $F_{CCS}$  and  $F_{SPLS}$ . Meanwhile, the selected structural and functional feature sets (i.e.,  $F_{GMVS}$ ,  $F_{CTS}$ ,  $F_{CCS}$  and  $F_{SPLS}$ ) for each training subject are used as inputs to train classifiers for MCI classification.

### 2.2.3. Multi-kernel SVM classification

Support vector machine (SVM) [49] is a classifier method based on kernel function, and has been widely used in various classification tasks. In this study we also use SVM as classifier. We first calculate four kernel matrices using the four selected feature sets (i.e.,  $F_{GMVS}$ ,  $F_{CTS}$ ,  $F_{CCS}$  or  $F_{SPLS}$ ) across different training subjects based on a linear kernel function as follows,

$$k(\mathbf{x}^i, \mathbf{x}^j) = (\mathbf{x}^i)' \mathbf{x}^j \quad (12)$$

where  $\mathbf{x}^i$  denotes the selected feature vector from the  $i$ th training subjects, and  $\mathbf{x}^j$  is similar to  $\mathbf{x}^i$ . Then, since multi-kernel SVM (MK-SVM) classification methods can effectively combine multiple features [24,50], in this study a MK-SVM classification method is also

adopted to combine these four different kernel matrices to train classifiers for MCI classification as shown in Fig. 1. The MK-SVM classification method can be formulated as follows,

$$k_M(\mathbf{x}^i, \mathbf{x}^j) = \sum_{m=1}^M \alpha_m k_m(\mathbf{x}_m^i, \mathbf{x}_m^j) \quad (13)$$

where  $M$  denotes the number of kernel matrices,  $k_m(\mathbf{x}_m^i, \mathbf{x}_m^j)$  denotes the  $m$ th kernel matrix that has been calculated, and  $\alpha_m$  denotes the combining weight on the  $m$ th kernel matrix. Obviously, in this study  $M = T = 4$ . Finally, a SVM classifier with  $k_M$  is used to perform the MCI classification task.

## 3. Experiments and results

### 3.1. Experimental settings

To investigate the performance of our proposed MCI classification method, three classification tasks in MCI classification are conducted as follows:

- T1: LMCI/NC classification, where LMCI is negative subjects and NC is positive subjects.
- T2: EMCI/NC classification, where EMCI is negative subjects and NC is positive subjects.
- T3: LMCI/EMCI classification, where LMCI is negative subjects and EMCI is positive subjects.

In this study, a nested 5-fold cross-validation strategy is adopted in the above three classification tasks. The nested 5-fold cross-validation mainly contains two aspects: an external 5-fold cross-validation and an internal 5-fold cross-validation. The external 5-fold cross-validation is to randomly divide all experimental subjects into 5 subsets without overlapping, one of which is left as the test set, and the remaining subsets are used for training. In each training procedure, an internal 5-fold cross-validation is adopted to determine the parameters (such as  $\lambda$ ,  $\beta$  and  $\alpha_m$ ) of our proposed MCI classification method. The internal 5-fold cross-validation is to randomly divide the training subjects into 5 subsets without overlapping, one of which is left as the validation set, and the remaining subsets are used for training. The purpose of each internal 5-fold cross-validation is to find the optimal parameters, which are used to generate an optimal classifier for each classification task via a grid search method. The optimal classifier for each classification task is used to distinguish the subjects in the corresponding testing set. To avoid the bias caused by randomly dividing the subjects in the cross-validation, the nested 5-fold cross-validation is repeated 50 times in T1, T2 and T3, respectively. In this study we only report the average of 50 repeated experiments in T1, T2 and T3, respectively.

In this study, the two parameters,  $\lambda$  and  $\beta$  in the feature selection procedure are set to  $[0, 100]$  with a step size of 5, and the parameters,  $\alpha_m$  in the classification procedure are set to  $[0, 1]$  with a step size of 0.1, and are constrained to  $\sum_{m=1}^M \alpha_m = 1$ . In addition, the SVM classifier in the classification procedure is implemented by using LIBSVM toolbox [51], and  $C = 1$ .

To quantitatively evaluate the performance of our proposed MCI classification method, in this study the three metrics are computed as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

$$SEN = \frac{TP}{TP + FN} \quad (15)$$

$$SPE = \frac{TN}{TN + FP} \quad (16)$$

where TP, TN, FP, and FN represent the numbers of true positive, true negative, false positive, and false negative, respectively;

**Table 3**  
Summary of the methods for comparison.

Data	Methods	Abbr.
T1w MRI	$F_{GMV} + \text{LASSO} + \text{L-SVM}$	M1
	$F_{CT} + \text{LASSO} + \text{L-SVM}$	M2
	$(F_{GMV} + F_{CT}) + \text{LASSO} + \text{L-SVM}$	M3
	$((F_{GMV} + \text{LASSO}) + (F_{CT} + \text{LASSO})) + \text{L-SVM}$	M4
	$((F_{GMV} + \text{LASSO}) \parallel (F_{CT} + \text{LASSO})) + \text{MK-SVM}$	M5
	$(F_{GMV} \parallel F_{CT}) + \text{MTFS-gLASSO} + \text{L-SVM}$	M6
	$(F_{GMV} \parallel F_{CT}) + \text{MTFS-gLASSO-TTR} + \text{L-SVM}$	M7
	$(F_{GMV} \parallel F_{CT}) + \text{MTFS-gLASSO} + \text{MK-SVM}$	M8
	$(F_{GMV} \parallel F_{CT}) + \text{MTFS-gLASSO-TTR} + \text{MK-SVM}$	M9
rs-fMRI	$F_{CC} + \text{LASSO} + \text{L-SVM}$	M10
	$F_{SPL} + \text{LASSO} + \text{L-SVM}$	M11
	$(F_{CC} + F_{SPL}) + \text{LASSO} + \text{L-SVM}$	M12
	$((F_{CC} + \text{LASSO}) + (F_{SPL} + \text{LASSO})) + \text{L-SVM}$	M13
	$((F_{CC} + \text{LASSO}) \parallel (F_{SPL} + \text{LASSO})) + \text{MK-SVM}$	M14
	$(F_{CC} \parallel F_{SPL}) + \text{MTFS-gLASSO} + \text{L-SVM}$	M15
	$(F_{CC} \parallel F_{SPL}) + \text{MTFS-gLASSO-TTR} + \text{L-SVM}$	M16
	$(F_{CC} \parallel F_{SPL}) + \text{MTFS-gLASSO} + \text{MK-SVM}$	M17
	$(F_{CC} \parallel F_{SPL}) + \text{MTFS-gLASSO-TTR} + \text{MK-SVM}$	M18
T1w	$(F_{GMVS} + F_{CTS} + F_{CCS} + F_{SPLS}) + \text{L-SVM}$	M19
MRI + rs-fMRI	$(F_{GMVS} \parallel F_{CTS} \parallel F_{CCS} \parallel F_{SPLS}) + \text{MK-SVM}$	M20

accuracy (ACC), sensitivity (SEN), and specificity (SPE) represent the ratio of subjects that are correctly identified, the ratio of positive subjects that are correctly identified, and the ratio of negative subjects that are correctly identified, respectively. In addition, to quantitatively evaluate the overall performance of our proposed MCI classification method, the area under receiver operating characteristic (ROC) curve (AUC) value [52] is also reported. It is worth mentioning that the greater the values of the four metrics are, the better the classification performance of the method is.

### 3.2. Methods for comparison

We conducted comparative experiments to validate the effectiveness of our proposed MCI classification method. Table 3 summarizes the methods under comparisons, which are denoted as M1–M20. Obviously, M20 is our proposed MCI classification method in this study. As can be seen from Table 3, these comparison methods mainly contain three categories: T1w MRI data-based methods (i.e., M1–M9), rs-fMRI data-based methods (i.e.,

M10–M18), and both T1w MRI and rs-fMRI data-based methods (i.e., M19–M20). These 20 methods are briefly introduced as follows.

For T1w MRI data-based methods, the process of M1 is that  $F_{GMV}$  is used as feature representation, the LASSO method is used to feature selection, and a linear SVM classifier (L-SVM) is applied to perform the MCI classification task. The only difference between M2 and M1 is that the feature representation of M2 is  $F_{CT}$ , and the only difference between M3 and M1 is that the feature representation of M3 is the concatenation of  $F_{GMV}$  and  $F_{CT}$ . The process of M4 is that concatenating the selected features from both M1 and M2 are used as feature representation, and a standard L-SVM classifier is used to perform the MCI classification task. In M5, we first calculate two kernels based on the features selected by M1 and the features based on M2 selection. Then the MK-SVM method is used to combine these two kernels to generate a new kernel, and train a SVM classifier with the new kernel to perform the MCI classification task. In M6, we first select features from  $F_{GMV}$  and  $F_{CT}$  using MTFS-gLASSO. Then the selected features are concatenated as input of a standard L-SVM classifier to perform the MCI classification task. The only difference between M7 and M6 is that the feature selection method of M7 is MTFS-gLASSO-TTR. In M8, we first select two feature subsets from  $F_{GMV}$  and  $F_{CT}$  using MTFS-gLASSO. Then the MK-SVM method is used to combine these two kernels to generate a new kernel, and train a SVM classifier with the new kernel to perform the MCI classification task. The only difference between M9 and M8 is that the feature selection method of M9 is MTFS-gLASSO-TTR.

For rs-fMRI data-based methods, these methods are similar with T1w MRI data-based methods. If  $F_{GMV}$  is replaced by  $F_{CC}$  and  $F_{CT}$  is replaced by  $F_{SPL}$ , M1–M9 and M10–M18 are in one-to-one correspondence.

For M19 and M20, the four feature subsets:  $F_{GMVS}$ ,  $F_{CTS}$ ,  $F_{CCS}$  and  $F_{SPLS}$  selected by using the MTFS-gLASSO-TTR method from  $F_{GMV}$ ,  $F_{CT}$ ,  $F_{CC}$  and  $F_{SPL}$ , respectively, are used as feature representation. The only difference between the two methods is that M19 is to concatenate the four selected feature subsets as input of a standard L-SVM classifier to perform the MCI classification task, while M20 is first to calculate four kernels based on these four selected feature subsets, and then a standard SVM classifier with the new kernel, which is a mixed kernel by combining these four kernels using MK-SVM, is used to perform the MCI classification task.

**Table 4**  
Results of 20 methods in three classification tasks, i.e., T1, T2 and T3, respectively.

Methods	T1				T2				T3			
	ACC(%)	SEN(%)	SPE(%)	AUC(%)	ACC(%)	SEN(%)	SPE(%)	AUC(%)	ACC(%)	SEN(%)	SPE(%)	AUC(%)
M1	69.9	72.4	67.5	70.1	64.8	67.5	62.4	65.8	58.9	57.8	60.5	60.1
M2	68.8	70.1	66.6	69.2	63.5	66.8	61.3	64.2	59.7	55.5	62.8	61.2
M3	72.4	68.5	73.7	72.5	66.9	64.8	69.6	67.2	61.6	60.2	63.4	62.8
M4	73.3	69.6	74.1	73.3	68.8	70.5	67.1	69.8	63.7	61.5	65.1	64.4
M5	74.5	75.2	71.1	74.8	70.4	69.5	73.3	71.2	64.6	62.2	65.7	65.6
M6	78.7	81.3	75.2	78.9	74.2	71.1	75.8	74.4	68.2	65.3	70.9	69.4
M7	80.7	76.8	85.4	80.4	75.1	73.9	76.5	75.5	70.5	68.6	72.8	71.1
M8	82.2	84.7	80.1	82.8	76.8	75.4	79.3	77.1	72.4	70.5	74.1	73.2
M9	83.7	80.5	85.1	83.3	78.1	<b>82.2</b>	75.1	78.8	73.5	70.9	75.8	74.4
M10	71.4	74.3	68.4	72.8	65.1	63.6	67.2	66.2	60.5	61.1	58.8	60.9
M11	72.5	69.3	74.8	73.4	65.9	64.9	67.3	66.8	61.7	63.3	60.7	62.5
M12	73.4	70.5	75.1	73.9	67.5	69.9	65.9	68.4	62.8	65.1	60.8	63.2
M13	73.9	77.5	69.5	74.1	69.1	65.2	72.8	69.5	64.2	67.5	61.8	64.8
M14	75.8	78.5	71.2	76.5	70.8	72.8	68.9	71.5	65.9	67.8	64.4	66.6
M15	79.2	76.4	80.7	79.9	74.5	77.8	71.8	75.1	71.5	75.9	68.3	72.3
M16	81.4	82.6	77.3	82.4	77.1	78.5	76.3	77.4	74.5	76.6	70.2	75.1
M17	84.5	85.5	81.7	84.8	77.3	75.6	80.8	77.8	75.4	77.5	72.3	75.9
M18	84.9	<b>87.7</b>	82.4	85.6	78.9	76.3	82.9	79.1	77.2	80.1	73.8	77.7
M19	85.4	82.9	88.7	86.7	80.3	78.2	82.1	79.6	78.5	80.3	75.3	78.9
M20	<b>88.5</b>	86.3	<b>90.3</b>	<b>89.7</b>	<b>82.7</b>	79.4	<b>83.9</b>	<b>83.2</b>	<b>79.6</b>	<b>83.8</b>	<b>76.8</b>	<b>80.3</b>

### 3.3. Classification performance

The classification performance of T1, T2 and T3 achieved by the above-mentioned 20 methods (i.e., M1–M20 in Table 3) are summarized in Table 4. From Table 4, we can observe that our proposed MCI classification method (i.e., M20) consistently outperforms the comparative methods in T1, T2 and T3, respectively. For example, our proposed MCI classification method achieves the ACC of 88.5%, 82.7% and 79.6% in T1, T2 and T3, respectively, while the best ACC achieved by other comparative methods is only 85.4%, 80.3% and 78.5% in T1, T2 and T3, respectively. Furthermore, our proposed MCI classification method achieves the AUC value of 0.897, 0.832 and 0.803 in T1, T2 and T3, respectively, while the best AUC value achieved by other comparative methods is only 0.867, 0.796 and 0.789 in T1, T2 and T3, respectively. These results show that our proposed method is efficient and advanced in MCI classification.

In addition, as can be seen from Table 4, the classification performance of the methods based on multiple feature sets (i.e., M3–M9 and M12–M20) is better than the classification performance of the methods based on single feature set alone (i.e., M1, M2, M10 and M11) in T1, T2 and T3, respectively. This result implies that different feature sets contain complementary information for MCI classification, and thus should be reasonably combined to improve the performance of MCI classification. Also, the classification performance of both T1w MRI and rs-fMRI data-based methods (i.e., M19–M20) is better than the classification performance of single-modal data-based (T1w MRI or rs-fMRI) methods (i.e., M1–M18) in T1, T2 and T3, respectively. This result shows that different modal data contain complementary information for MCI classification, and thus should also be reasonably combined to improve the performance of MCI classification. Furthermore, the classification performance of the methods based on regional functional features (i.e., M10–M18) is better than the classification performance of the methods based on regional structural features (i.e., M1–M9) in T1, T2 and T3, respectively. To some extent, this result shows that regional functional features can better express differences than regional structural features for MCI classification.

## 4. Discussion

### 4.1. Effects of regularization parameters

In this study, we propose an improved multi-task feature selection method, namely MTFs-gLASSO-TTR, to select discriminative features for MCI classification. The MTFs-gLASSO-TTR method contains two regularization items, i.e., gLASSO regularizer and TTR regularizer, which are balanced by two regularization parameters,  $\lambda$  and  $\beta$ . To investigate the effects of the two regularization parameters,  $\lambda$  and  $\beta$  for MCI classification, we select different values of  $\lambda$  and  $\beta$  to perform three different classification tasks, i.e., T1, T2 and T3. In the three different classification tasks, the two regularization parameters,  $\lambda$  and  $\beta$  are set to [0, 100] with a step size of 5. It is worth mentioning that when  $\lambda = 0$ ,  $\beta = 0$ , that is, our proposed MCI classification method does not include the feature selection step; when  $\lambda > 0$  and  $\beta = 0$ , the feature selection step of our proposed MCI classification method is the MTFs-gLASSO-based method. Fig. 3 shows the classification accuracies based on different values of  $\lambda$  and  $\beta$  in T1, T2 and T3.

From Fig. 3, we can see that the classification accuracies of the methods with feature selection (i.e.,  $\lambda > 0$ ) are better than that of the method without feature selection (i.e.,  $\lambda = 0$ ) in T1, T2 and T3, respectively. This result implies that the feature selection step can obtain more discriminative features for MCI classification. Also, the classification accuracies of the method with TTR regularizer (i.e.,  $\beta > 0$ ) are better than that of the method without TTR regularizer

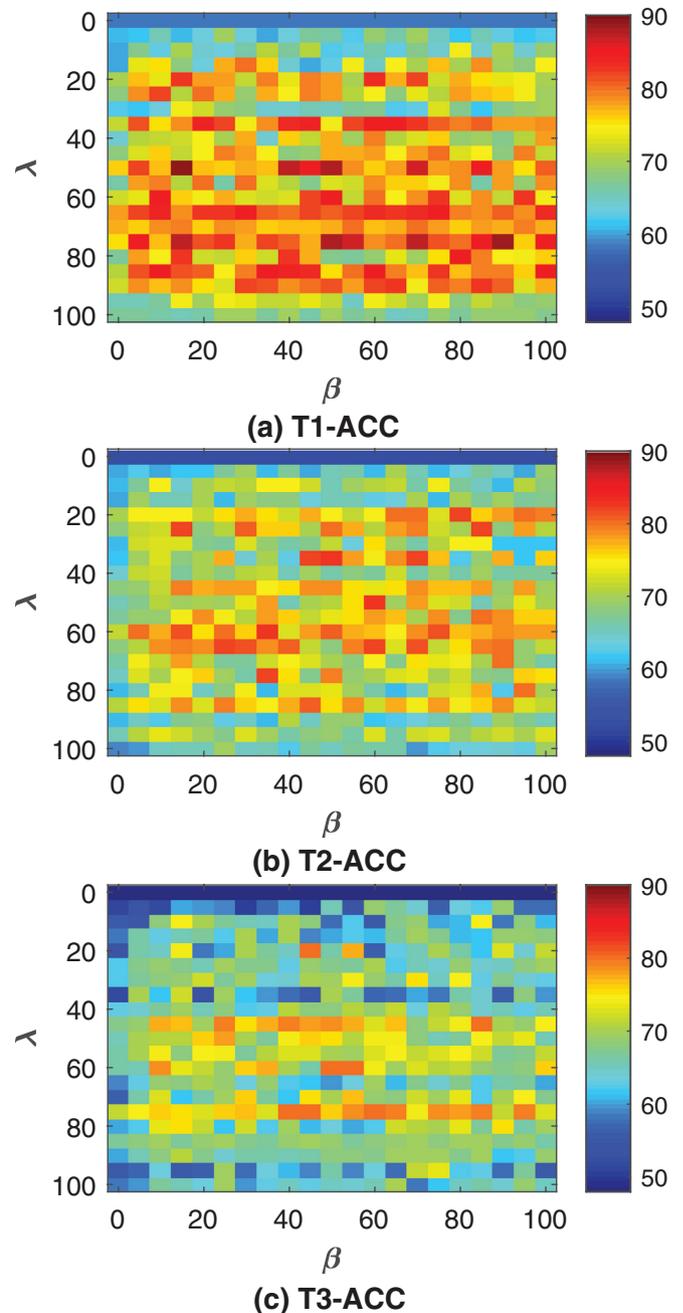
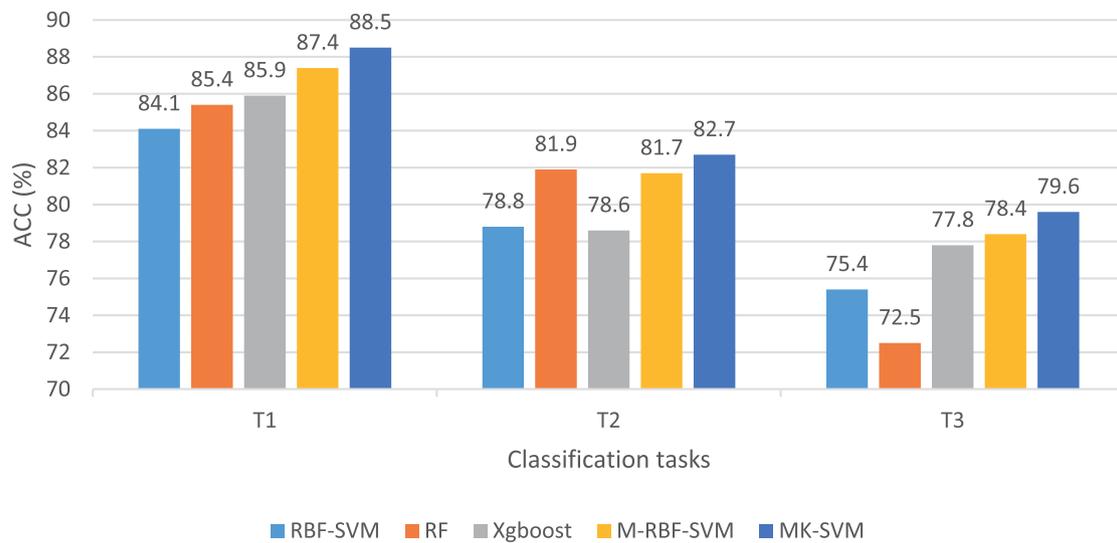


Fig. 3. Results achieved by our proposed MCI classification method (i.e., M20) based on different values of  $\lambda$  and  $\beta$  in T1, T2 and T3, respectively.

(i.e.,  $\beta = 0$ ) in T1, T2 and T3, respectively. This result suggests that adding TTR regularizer into the MTFs-gLASSO-based method can obtain more discriminative features for MCI classification. In addition, we can also see that the classification accuracies are greatly affected by different values of  $\lambda$  and  $\beta$  in T1, T2 and T3, respectively. This result indicates that it is very important to balance the two regularization parameters,  $\lambda$  and  $\beta$  for MCI classification.

### 4.2. Effects of different classifiers

In this subsection, we investigate the effects of different classifiers in T1, T2 and T3, respectively. For this purpose, we have done a series of experiments using different classifiers including radial basis function kernel-support vector machine (RBF-SVM), random forest (RF) [53], Xgboost [54] and multiple RBF-SVM (M-RBF-SVM)



**Fig. 4.** The accuracy of our proposed method using different classifiers for T1, T2 and T3, respectively.

**Table 5**

The best ACC of MCI classification based on different combination weights in T1, T2 and T3, respectively.

Tasks	Q1	Q2	Q3	Q4
T1	<b>88.5</b>	86.3	85.2	73.6
T2	<b>82.7</b>	80.8	79.5	67.7
T3	<b>79.6</b>	78.1	77.6	64.3

in our proposed method. The input of the first three classifiers is a concatenation of the four selected feature sets obtained after feature selection for each classification task. The last classifier is similar to our proposed MK-SVM method, and the only difference between them is that the RBF is used instead of the linear kernel. It is worth mentioning that these classifiers with the default parameters are implemented by the Scikit-learn library [55]. In addition, for fair comparisons, the experimental data used by each classifier is the same for the same classification task. The accuracy of our proposed method using different classifiers is shown in Fig. 4 for T1, T2 and T3, respectively.

As can be seen from Fig. 4, the ACC of MK-SVM in our proposed method is the best for T1, T2 and T3, respectively. The comparative results further demonstrate that our proposed method is effective and has certain advantages for MCI classification.

#### 4.3. Effects of different combination weights

In this subsection, we investigate the effects of different combination weights  $\alpha_m$ ,  $m \in \{1, 2, 3, 4\}$  in T1, T2 and T3, respectively. Since the values of  $\alpha_m$  are  $[0, 1]$  with a step size of 0.1, and are constrained to  $\sum_{m=1}^M \alpha_m = 1$ , there are many combination of  $\alpha_m$ . In this subsection, we only report the best ACC of MCI classification when none of the  $\alpha_m$  values is 0 (denoted as Q1), one of the  $\alpha_m$  values is 0 (denoted as Q2), two of the  $\alpha_m$  values are 0 (denoted as Q3), and three of the  $\alpha_m$  values are 0 (denoted as Q4) in each combination in T1, T2 and T3, respectively. Table 5 shows the best ACC of MCI classification based on different combination weights in T1, T2 and T3, respectively.

As can be seen from Table 5, the best ACC of Q1 is better than that of the other three cases (i.e., Q2, Q3 and Q4) in T1, T2 and T3, respectively. This result indicates that our proposed MCI classification method is effective. In addition, the best ACC of Q4 is better than that of single feature set-based classification method (i.e., M1, M2, M10 and M11) in T1, T2 and T3, respectively. This result

implies that compared with STFS method, for a specific original feature set, MTFS method can select more discriminant features in T1, T2 and T3, respectively, which may be due to considering the relationship between tasks.

#### 4.4. Important regional features

In this subsection, we report the important regional features extracted from both T1w MRI and rs-fMRI data by our proposed MCI classification method in T1, T2 and T3, respectively. In general, the features used for training classifier after feature selection in each fold cross-validation are different in a specific classification task. Therefore, in this study we define the features that appear in the features used for training classifier in each fold cross-validation as important regional features in each MCI classification task. Tables 6, and 8 show the important regional features obtained by our proposed MCI classification method in T1, T2 and T3, respectively. Furthermore, we also report  $p$ -value of each important regional feature via a standard two-sample  $t$ -test to reveal their discriminative power in T1, T2 and T3, respectively.

As can be seen from Tables 6, and 8, most important brain regions, such as PHG, INS and MTP, obtained by our proposed MCI classification method have been reported in previous MCI studies [21,38,56]. This result indicates that our proposed MCI classification method can detect brain regions related to MCI, and is effective for MCI classification. Meanwhile, we also can see that important regional features contain both structural and functional features. This result implies that both structural and functional features are important for MCI classification, and MCI is associated with brain structural and functional changes. Furthermore, there are more important functional features than important structural features in T1, T2 and T3, respectively. This result explains to some extent that the classification performance of the methods only using functional features is better than the classification performance of the methods only using structural features, which is shown in Table 4. In addition, most important regional features have a  $p$ -value of less than 0.05. This result indicates that the important brain features obtained by our proposed MCI classification method have good discriminative power for MCI classification.

As can be seen from Tables 6 and 7, more important regional features are involved in T1, compared with those in T2. This result indicates that more brain region changes are produced with the development of disease progression (i.e., NC  $\rightarrow$  EMCI  $\rightarrow$  LMCI).

**Table 6**  
Important regional features involved in T1.L = Left.,R = Right.

No.	Feature types	Cortical regions	p-values
1	GMV	INS.L	<0.001
2	GMV	PrCG.R	0.009
3	GMV	MTP.R	0.015
4	GMV	MOC.R	0.018
5	GMV	PHG.L	0.024
6	GMV	LING.R	0.029
7	GMV	INS.L	0.034
8	GMV	STG.R	0.048
9	GMV	SOG.L	0.086
10	CT	CUN.R	<0.001
11	CT	IOG.R	0.002
12	CT	INS.L	0.012
13	CT	REC.R	0.014
14	CT	ANG.L	0.036
15	CT	MTP.R	0.047
16	CT	PCUN.L	0.061
17	CC	IPL.L	<0.001
18	CC	CAL.L	<0.001
19	CC	CUN.R	0.008
20	CC	PHG.L	0.013
21	CC	SOG.R	0.017
22	CC	PCL.R	0.023
23	CC	REC.R	0.029
24	CC	INS.R	0.039
26	CC	PrCG.L	0.042
27	CC	STP.L	0.053
28	CC	LING.R	0.077
29	SPL	PHG.L	<0.001
30	SPL	LING.R	0.014
31	SPL	REC.L	0.022
32	SPL	PrCG.L	0.026
33	SPL	IOG.L	0.038
34	SPL	PCL.R	0.042
36	SPL	INS.L	0.078
37	SPL	ROL.R	0.094

**Table 7**  
Important regional features involved in T2.L = Left.,R = Right.

No.	Feature types	Cortical regions	p-values
1	GMV	PHG.L	0.031
2	GMV	PrCG.R	0.038
3	GMV	INS.L	0.041
4	GMV	MTP.R	0.046
5	GMV	STG.R	0.055
6	CT	REC.R	0.022
7	CT	INS.L	0.028
8	CT	CUN.L	0.037
9	CT	ANG.L	0.048
10	CC	CUN.R	0.017
11	CC	MTP.R	0.023
12	CC	PHG.L	0.025
13	CC	IPL.L	0.032
15	CC	PoCG.R	0.041
16	CC	CAL.L	0.048
17	SPL	PHG.L	0.015
18	SPL	REC.L	0.031
19	SPL	MTP.L	0.034
20	SPL	LING.R	0.047
21	SPL	IPL.R	0.061

Meanwhile, as can be seen from Tables 7 and 8, less important regional features are involved in T3, compared with those in T2. This result indicates that the early stage of disease progression is faster, but the medium term is relatively slow with the development of disease progression (i.e., NC → EMCI → LMCI). Such results indicate that early classification of MCI is very important for human health.

**Table 8**  
Important regional features involved in T3.L = Left.,R = Right.

No.	Feature types	Cortical regions	p-values
1	GMV	PoCG.R	0.018
2	GMV	STP.R	0.047
3	CT	MCG.L	0.032
4	CT	PHG.R	0.038
5	CT	ANG.L	0.042
6	CC	ITG.R	0.015
7	CC	MeFG.L	0.027
8	CC	IFGorb.R	0.033
9	CC	SFG.L	0.039
10	SPL	REC.L	0.035
11	SPL	PCG.R	0.041
12	SPL	ROL.L	0.048

**Table 9**  
Comparison with four state-of-the-art methods in T1, T2 and T3, respectively.

Tasks	Methods	ACC(%)	SEN(%)	SPE(%)	AUC(%)	p-values
T1	Sarraf and Tofighi [33]	82.9	86.1	79.9	84.2	< 0.05
	Tripathi et al. [23]	85.9	83.4	87.9	86.3	< 0.05
	Islam and Zhang [32]	87.1	85.2	88.9	88.3	< 0.05
	Jie et al. [24]	86.8	84.5	89.7	87.8	< 0.05
	Our proposed	<b>88.5</b>	<b>86.3</b>	<b>90.3</b>	<b>89.7</b>	
T2	Sarraf and Tofighi [33]	77.4	78.8	75.6	78.6	< 0.05
	Tripathi et al. [23]	75.8	74.2	76.7	76.2	< 0.05
	Islam and Zhang [32]	81.1	77.5	83.7	81.9	< 0.05
	Jie et al. [24]	79.5	<b>82.6</b>	77.2	80.1	< 0.05
	Our proposed	<b>82.7</b>	79.4	<b>83.9</b>	<b>83.2</b>	
T3	Sarraf and Tofighi [33]	73.7	77.2	70.5	74.4	< 0.05
	Tripathi et al. [23]	71.9	77.6	66.7	72.7	< 0.05
	Islam and Zhang [32]	76.9	80.7	75.3	77.7	< 0.05
	Jie et al. [24]	78.1	74.7	<b>80.9</b>	78.5	< 0.05
	Our proposed	<b>79.6</b>	<b>83.8</b>	76.8	<b>80.3</b>	

#### 4.5. Comparisons with state-of-the-art methods

To demonstrate the superiority of our proposed MCI classification method, we also compare four existing state-of-the-art methods in the field including traditional machine learning methods [23,24] and deep learning methods [32,33]. In our comparative experiments, all methods are repeated 50 times for 5-fold cross-validation, and the average classification performance in T1, T2 and T3, respectively, are reported in Table 9. In order to statistically verify that the performance of our proposed MCI classification method is better than that of other four existing state-of-the-art methods, we also report the p-values of the proposed method and other methods in terms of ACC, which is shown in Table 9.

From Table 9, we can see that our proposed MCI classification method obtains the best classification ACC and AUC values in T1, T2 and T3, respectively. Although Jie et al. [24] method obtains the best SEN in T2, and the best SPE in T3, the other three metrics of this method in T2 and T3 are lower than those of our proposed method, respectively. In addition, as the statistical p-value is less than 0.05, our proposed MCI classification method is significantly better than other four existing state-of-the-art methods.

#### 4.6. Limitations

While addressing some challenges, we have also identified a number of limitations in our proposed method. Firstly, the subjects used in our work are only a portion of the ADNI database, which may not represent the pathological characteristics of large MCI populations. Applying our proposed method to a larger sample size might reveal more convincing results. Secondly, as the brain atlas is an influential factor to classification accuracy, more proper partitioning would probably lead to better results. As the future

work, we will test the classification performance of our proposed MCI classification method based on different atlas. In addition, our proposed method uses only two structural features and two functional features. However, using more features may obtain better classification accuracy. Therefore, future work will try to combine more structural and functional features for our proposed method.

## 5. Conclusion

In this study, we propose a new method to enhance the feature representation of multi-modal MRI data by combining multi-view information to improve the performance of MCI classification. Firstly, we perform image preprocessing and feature representation for both T1w MRI and rs-fMRI data of each subject. After this step, we obtain four feature sets including two structural feature sets and two functional feature sets for each subject. Then, an improved multi-task feature selection method, namely MTFs-gLASSO-TTR, is proposed to select the optimal structural and functional features from these four feature sets for MCI classification. Finally, a MK-SVM classification method is adopted to combine these four selected structural and functional feature sets to perform the MCI classification task. Experimental results on 315 subjects from ADNI database demonstrate that our proposed method is effective in MCI classification. This method paves the way to discriminative imaging markers for computer-aided classification of MCI.

## Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interests.

## CRediT authorship contribution statement

**Jin Liu:** Conceptualization, Methodology, Software, Writing - review & editing. **Yi Pan:** Methodology. **Fang-Xiang Wu:** Methodology. **Jianxin Wang:** Conceptualization, Methodology, Writing - review & editing.

## Acknowledgments

This work is supported in part by the [National Natural Science Foundation of China](#) under Grant Nos. 61802442 and 61877059, the [Natural Science Foundation of Hunan Province](#) under Grant No. 2019JJ50775, the 111 Project (No. B18059), and the [Hunan Provincial Science and Technology Program](#) (2018WK4001).

## References

- [1] A. Association, 2016 Alzheimer's disease facts and figures, *Alzheimers Dement.* 12 (4) (2016) 459–509.
- [2] K. Kantarci, S.D. Weigand, S.A. Przybelski, M.M. Shiung, J.L. Whitwell, S. Negash, D.S. Knopman, B.F. Boeve, P.C. O'Brien, R.C. Petersen, Risk of dementia in MCI: combined effect of cerebrovascular disease, volumetric MRI, and 1h MRS, *Neurology* 72 (17) (2009) 1519–1525.
- [3] J. Liu, J. Wang, H. Bin, F.-X. Wu, Y. Pan, Alzheimer's disease classification based on individual hierarchical networks constructed with 3D texture features, *IEEE Trans. Nanobiosci.* 16 (6) (2017) 428–437.
- [4] Y. Li, J. Liu, J. Huang, Z. Li, P. Liang, Learning brain connectivity sub-networks by group-constrained sparse inverse covariance estimation for Alzheimer's disease classification, *Front. Neuroinform.* 12 (58) (2018) 1–13.
- [5] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, J. Wang, Classification of Alzheimer's disease using whole brain hierarchical network, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15 (2) (2018) 624–632.
- [6] Y. Li, H. Yang, K. Li, P.T. Yap, M. Kim, C.Y. Wee, D. Shen, Novel effective connectivity network inference for MCI identification, in: *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 316–324.
- [7] J. Liu, M. Li, Y. Pan, F.-X. Wu, X. Chen, J. Wang, Classification of schizophrenia based on individual hierarchical brain networks constructed from structural MRI images, *IEEE Trans. Nanobiosci.* 16 (7) (2017) 600–608.
- [8] Y. Xiang, J. Wang, G. Tan, F.-X. Wu, J. Liu, Schizophrenia identification using multi-view graph measures of functional brain networks, *Front. Bioeng. Biotechnol.* 7 (2020) 479.
- [9] Y. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang, J. Liu, Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier, *Neurocomputing* 324 (9) (2019) 63–68.
- [10] J. Wang, Q. Wang, J. Peng, D. Nie, F. Zhao, M. Kim, H. Zhang, C.-Y. Wee, S. Wang, D. Shen, Multi-task diagnosis for autism spectrum disorders using multi-modality features: a multi-center study, *Hum. Brain Mapp.* 38 (6) (2017) 3081–3097.
- [11] G. Karas, P. Scheltens, S. Rombouts, P. Visser, R. Van Schijndel, N. Fox, F. Barkhof, Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease, *Neuroimage* 23 (2) (2004) 708–716.
- [12] J.H. Morra, Z. Tu, L.G. Apostolova, A.E. Green, C. Avedissian, S.K. Madsen, N. Parikshak, X. Hua, A.W. Toga, C.R. Jack, et al., Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls, *Neuroimage* 43 (1) (2008) 59–68.
- [13] M. Khajehnejad, F.H. Saatlou, H. Mohammadzade, Alzheimer's disease early diagnosis using manifold-based semi-supervised learning, *Brain Sci.* 7 (8) (2017) 109.
- [14] J. Liu, X. Wang, X. Zhang, Y. Pan, X. Wang, J. Wang, MMM: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning, *Multimed. Tools Appl.* 77 (22) (2018) 29651–29667.
- [15] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-P-SO-optimized SVM algorithm for diagnosis of Alzheimer's disease, *Neurocomputing* 320 (2018) 195–202.
- [16] G. Chen, B.D. Ward, C. Xie, W. Li, Z. Wu, J.L. Jones, M. Franczak, P. Antuono, S.-J. Li, Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging, *Radiology* 259 (1) (2011) 213–221.
- [17] Y. Feng, L. Bai, Y. Ren, S. Chen, H. Wang, W. Zhang, J. Tian, Fmri connectivity analysis of acupuncture effects on the whole brain network in mild cognitive impairment patients, *Magn. Reson. Imaging* 30 (5) (2012) 672–682.
- [18] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao, J. Jia, Y. Han, Y. He, Disrupted functional brain connectome in individuals at risk for Alzheimer's disease, *Biol. Psychiatry* 73 (5) (2013) 472–481.
- [19] M.R. Brier, J.B. Thomas, A.M. Fagan, H. Jason, D.M. Holtzman, T.L. Benzinger, J.C. Morris, B.M. Ances, Functional connectivity and graph theory in preclinical Alzheimer's disease, *Neurobiol. Aging* 35 (4) (2014) 757–768.
- [20] X. Zhang, B. Hu, X. Ma, L. Xu, Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression, *IEEE Trans. Nanobiosci.* 14 (2) (2015) 237–247.
- [21] C.-Y. Wee, P.-T. Yap, D. Zhang, K. Denny, J.N. Browndyke, G.G. Potter, K.A. Welsh-Bohmer, L. Wang, D. Shen, Identification of MCI individuals using structural and functional connectivity networks, *Neuroimage* 59 (3) (2012) 2045–2056.
- [22] M. De Marco, L. Beltrachini, A. Biancardi, A.F. Frangi, A. Venneri, Machine-learning support to individual diagnosis of mild cognitive impairment using multimodal MRI and cognitive assessments, *Alzheimer Dis. Assoc. Disord.* 31 (4) (2017) 278–286.
- [23] S. Tripathi, S.H. Nozadi, M. Shakeri, S. Kadoury, Sub-cortical shape morphology and voxel-based features for Alzheimer's disease classification, in: *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2017, pp. 991–994.
- [24] B. Jie, M. Liu, D. Shen, Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease, *Med. Image Anal.* 47 (2018) 81–94.
- [25] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, T. Young, Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach, *IEEE Trans. Med. Imaging* 33 (5) (2014) 1129–1136.
- [26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [27] Y. Yu, M. Li, L. Liu, Z. Fei, F.-X. Wu, J. Wang, Automatic ICD code assignment of chinese clinical notes based on multilayer attention Birn, *J. Biomed. Inform.* 91 (2019) 103114.
- [28] Y. An, N. Huang, X. Chen, F. Wu, J. Wang, High-risk prediction of cardiovascular diseases via attention-based deep neural networks, *IEEE/ACM Trans. Comput. Biol. Bioinform.* PP (99) (2019) 1–1.
- [29] N. Zeng, Z. Wang, H. Zhang, K.-E. Kim, Y. Li, X. Liu, An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips, *IEEE Trans. Nanotechnol.* 18 (2019) 819–829.
- [30] L. Liu, F.-X. Wu, J. Wang, Efficient multi-kernel DCNN with pixel dropout for stroke MRI segmentation, *Neurocomputing* 350 (2019) 117–127.
- [31] L. Liu, S. Chen, X. Zhu, X.-M. Zhao, F.-X. Wu, J. Wang, Deep convolutional neural network for accurate segmentation and quantification of white matter hyperintensities, *Neurocomputing* 384 (2020) 231–242.
- [32] J. Islam, Y. Zhang, Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks, *Brain Inform.* 5 (2) (2018) 2.
- [33] S. Sarraf, G. Tofghi, Classification of Alzheimer's Disease Using fMRI Data and Deep Learning Convolutional Neural Networks, *arXiv preprint:1603.08631* (2016).
- [34] S. Sarraf, G. Tofghi, Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, in: *Proceedings of the 2016 Future Technologies Conference (FTC)*, IEEE, 2016, pp. 816–820.
- [35] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in

- SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, *Neuroimage* 15 (1) (2002) 273–289.
- [36] M.C. Carrillo, L.J. Bain, G.B. Frisoni, M.W. Weiner, Worldwide Alzheimer's disease neuroimaging initiative, *Alzheimer's Dement.* 8 (4) (2012) 337–342.
- [37] B. Fischl, Freesurfer, *Neuroimage* 62 (2) (2012) 774–781.
- [38] W. Zheng, Z. Yao, B. Hu, X. Gao, H. Cai, P. Moore, Novel cortical thickness pattern for accurate detection of Alzheimer's disease, *J. Alzheimers Dis.* 48 (4) (2015) 995–1008.
- [39] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, Y. Pan, Improving Alzheimer's disease classification by combining multiple measures, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 15 (5) (2018) 1649–1659.
- [40] R.W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages, *Comput. Biomed. Res.* 29 (3) (1996) 162.
- [41] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: uses and interpretations, *Neuroimage* 52 (3) (2010) 1059–1069.
- [42] D.S. Bassett, O. Sporns, Network neuroscience, *Nat. Neurosci.* 20 (3) (2017) 353–364.
- [43] J. Liu, M. Li, Y. Pan, W. Lan, R. Zheng, F.-X. Wu, J. Wang, Complex brain network analysis and its applications to brain disorders: a survey, *Complexity* (2017), doi:10.1155/2017/8362741.
- [44] Z. Yao, Y. Zhang, L. Lin, Y. Zhou, C. Xu, T. Jiang, A.D.N. Initiative, et al., Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease, *PLoS Comput. Biol.* 6 (11) (2010) e1001006.
- [45] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393 (6684) (1998) 440–442.
- [46] J.-P. Onnela, J. Saramaki, J. Kertész, K. Kaski, Intensity and coherence of motifs in weighted complex networks, *Phys. Rev. E* 71 (6) (2005) 065103.
- [47] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B (Methodol.)* 58 (1) (1996) 267–288.
- [48] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc.* 68 (1) (2006) 49–67.
- [49] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297.
- [50] A. Rakotomamonjy, F. Bach, S. Canu, Y. Grandvalet, More efficiency in multiple kernel learning, in: *Proceedings of the 24th International Conference on Machine Learning, ACM, 2007*, pp. 775–782.
- [51] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol. (TIST)* 2 (3) (2011) 1–27.
- [52] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology* 143 (1) (1982) 29–36.
- [53] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, *R News* 2 (3) (2002) 18–22.
- [54] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016*, pp. 785–794.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: machine learning in python, *J. Mach. Learn. Res.* 12 (Oct) (2011) 2825–2830.
- [56] H. Wolf, V. Jelic, H.-J. Gertz, A. Nordberg, P. Julin, L.-O. Wahlund, A critical discussion of the role of neuroimaging in mild cognitive impairment, *Acta Neurol. Scand.* 107 (s179) (2003) 52–76.



**Jin Liu** received the Ph.D. degree in computer science from Central South University, China, in 2017. He is currently a lecturer at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis, bioinformatics and machine learning.



**Yi Pan** is a Regents' Professor of Computer Science and an Interim Associate Dean and Chair of Biology at Georgia State University, USA. He joined Georgia State University in 2000 and was promoted to full professor in 2004, named a Distinguished University Professor in 2013 and designated a Regents' Professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. He served as the Chair of Computer Science Department from 2005 to 2013. He is also a visiting Changjiang Chair Professor at Central South University, China. He received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pittsburgh, USA, in 1991. His profile has been featured as a distinguished alumnus in both Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. His research interests include parallel and cloud computing, wireless networks, and bioinformatics. He has published more than 330 papers including over 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 8800 times. He has served as an editor-in-chief or editorial board member for 15 journals including 7 IEEE Transactions. He is the recipient of many awards including IEEE Transactions Best Paper Award, 4 other international conference or journal Best Paper Awards, 4 IBM Faculty Awards, 2 JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.



**Fang-Xiang Wu** received the B.Sc. degree and the M.Sc. degree in applied mathematics, both from Dalian University of Technology, Dalian, China, in 1990 and 1993, respectively, the first Ph.D. degree in control theory and its applications from Northwestern Polytechnical University, Xi'an, China, in 1998, and the second Ph.D. degree in biomedical engineering from University of Saskatchewan (U of S), Saskatoon, Canada, in 2004. During 2004–2005, he worked as a Postdoctoral Fellow in the Laval University Medical Research Center (CHUL), Quebec City, Canada. He is currently a Professor of the Division of Biomedical Engineering and the Department of Mechanical Engineering at the U of S. His current research interests include computational and systems biology, genomic and proteomic data analysis, biological system identification and parameter estimation, applications of control theory to biological systems. He has published more than 260 technical papers in refereed journals and conference proceedings. He is serving as the editorial board member of three international journals, the guest editor of several international journals, and as the program committee chair or member of several international conferences. He has also reviewed papers for many international journals.



**Jianxin Wang** received the B.Eng. and M.Eng. degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the Ph.D. degree in computer science from Central South University, China, in 2001. He is the chair of and a professor in School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, Bioinformatics and computer network.