Improving Alzheimer's Disease Classification by Combining Multiple Measures

Jin Liu, Jianxin Wang, Zhenjun Tang, Bin Hu, Fang-Xiang Wu, and Yi Pan

Abstract—Several anatomical magnetic resonance imaging (MRI) markers for Alzheimer's disease (AD) have been identified. Cortical gray matter volume, cortical thickness, and subcortical volume have been used successfully to assist the diagnosis of Alzheimer's disease including its early warning and developing stages, e.g., mild cognitive impairment (MCI) including MCI converted to AD (MCIc) and MCI not converted to AD (MCInc). Currently, these anatomical MRI measures have mainly been used separately. Thus, the full potential of anatomical MRI scans for AD diagnosis might not yet have been used optimally. Meanwhile, most studies currently only focused on morphological features of regions of interest (ROIs) or interregional features without considering the combination of them. To further improve the diagnosis of AD, we propose a novel approach of extracting ROI features and interregional features based on multiple measures from MRI images to distinguish AD, MCI (including MCIc and MCInc), and health control (HC). First, we construct six individual networks based on six different anatomical measures (i.e., CGMV, CT, CSA, CC, CFI, and SV) and Automated Anatomical Labeling (AAL) atlas for each subject. Then, for each individual network, we extract all node (ROI) features and edge (interregional) features, and denoted as node feature set and edge feature set, respectively. Therefore, we can obtain six node feature sets and six edge feature sets from six different anatomical measures. Next, each feature within a feature set is ranked by F-score in descending order, and the top ranked features of each feature set are applied to MKBoost algorithm to obtain the best classification accuracy. After obtaining the best classification accuracy, we can get the optimal feature subset and the corresponding classifier for each node or edge feature set. Afterwards, to investigate the classification performance with only node features, we proposed a weighted multiple kernel learning (wMKL) framework to combine these six optimal node feature subsets, and obtain a combined classifier to perform AD classification. Similarly, we can obtain the classification performance with only edge features. Finally, we combine both six optimal node feature subsets and six optimal edge feature subsets to further improve the classification performance. Experimental results show that the proposed method outperforms some state-of-the-art methods in AD classification, and demonstrate that different measures contain complementary information.

Index Terms—Alzheimer's disease, anatomical MRI, multiple measures, multiple kernel learning, classification

1 Introduction

Alzheimer's Disease (AD) is the most common neurodegenerative disorder among aging people, which accounts for nearly 70 percent of all dementia cases [1]. The early stages of AD include a noticeable and measurable decline in memory, language, thinking and other cognitive abilities. Patients with these symptoms are usually diagnosed as the Mild Cognitive Impairment (MCI) [2]. MCI does not notably interfere with daily activities, but those with MCI have a higher risk of later progression to AD or other forms of dementia [3]. Many medical interventions may only be effective in the early course of the disease [4]. Therefore, precise prediction and diagnosis of AD, especially the detection of MCI, could help the physicians to identify patients with a high risk of developing dementia and allow the patients to receive early medical interventions before irreversible brain damages are formed.

Neuroimaging technology has been widely used in disease research because it is able to non-invasively measure the structural and functional changes associated with the development of diseases *in vivo*, such as brain tumor [5], [6], schizophrenia [7], [8], AD [9], [10] and so on. The earliest technique for brain structure imaging is computed tomography. Currently, computed tomography has been largely replaced by magnetic resonance imaging (MRI) because MRI can produce high quality three-dimensional (3D) images of brain structures by using magnetic fields and radio waves instead of ionizing radiation or radioactive tracers [11]. MRI is a more powerful and safe technique and provides more detailed structural information. MRI falls into two broad categories: structural (anatomical) imaging and functional imaging. Owning to its easy access in clinical settings, structural MRI receives more attention of

- *J. Liu and J. Wang are with the School of Information Science and Engineering, Central South University, Changsha 410083, China. E-mail: {liujin06, jxwang}@mail.csu.edu.cn.*
- *Z. Tang is with the School of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China. E-mail: zjtang@gxnu.edu.cn.*
- *B. Hu is with the School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China. E-mail: bh@lzu.edu.cn.*
- *F.X. Wu is with the Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon SK S7N 5A9, Canada. E-mail: faw341@mail.usask.ca.*
- *Y. Pan is with the Department of Computer Science, Georgia State University, Atlanta, GA 30302. E-mail: yipan@gsu.edu.*

Manuscript received 30 Nov. 2016; revised 19 June 2017; accepted 21 July 2017. Date of publication 25 July 2017; date of current version 5 Oct. 2018. (Corresponding author: Jianxin Wang.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2017.2731849

researchers compared with the other. T_1-weighted MRI images, a common type of structural MRI, are used in our study.

Region-based analysis methods have become one of the most popular approaches for automatic AD diagnosis. The key of region-based analysis methods is the determination of the region of interest (ROI). Once ROIs in the brain are determined, they can be utilized to identify the anatomical differences between populations of AD, MCI and health control (HC), and subsequently determine the AD-related characteristics to assist diagnosis, prognosis, as well as evaluation of MCI progression and treatment effects. Region-based analysis methods can be divided into two categories according to the number of ROIs: single ROI method [12], [13] and multiple ROIs method [9], [14], [15]. For example, Chupin et al. [13] proposed to use the hippocampal volumes to automatically discriminate between populations of AD, MCI and HC. Ahmed et al. [14] proposed an automatic classification framework based on the fusion of both hippocampus visual features and cerebrospinal fluid volume for AD recognition in structural MRI. Furthermore, some researchers believe that AD is related to the whole brain. For example, to improve AD classification performance, Linn et al. [15] proposed to parcellate the brain of each subject into 137 ROIs, and Liu et al. [9] proposed to parcellate the brain of each subject into 90 ROIs, 54 ROIs, 14 ROIs and 1 ROI. In [9], Liu et al. demonstrated that AD is related to not only ROIs, but also the correlation between two ROIs. Therefore, in this article, we consider not only each ROI in the brain, but also the correlation between two ROIs.

In the literature, there are a number of metrics to measure the anatomical changes associated with AD by using brain MRI images, such as cortical gray matter volume (CGMV), cortical thickness (CT), cortical surface area (CSA), cortical curvature (CC), cortical folding index (CFI), subcortical volume (SV) and so on. Furthermore, these anatomical MRI measures have been widely used in the study of AD. For example, Karas et al. [16] concluded that the loss of CGMV in the medial temporal lobe (MTL) characterized MCI, while the loss of CGMV in the parietal and cingulate cortices might be a feature of AD. Singh et al. [17] considered CT as a more stable parameter for AD diagnosis than volume measures, because it is a more direct measure of GM atrophy due to the cytoarchitectural feature of the GM. Bobinski et al. [18] indicated that the measurement of the CSA of the entorhinal cortex (EC) in MRI images was potentially useful in the early diagnosis of AD. Morra et al. [19] considered CC as a measure method to study the difference among AD, MCI and HC. Cash et al. [20] used the CFI as a biomarker of AD and indicated that a significant reduction was observed between HC and AD in some ROIs associated with the sulcal folds.

Thus far, these anatomical MRI measures have mainly been used separately for automatic AD diagnosis. However, some researchers have also used the combination of different structural MRI measures to improve AD classification [21], [22]. For example, Westman et al. [21] gave a good prediction accuracy by combining regional CT and SV using orthogonal partial least square (OPLS) method when distinguishing AD versus HC. Therefore, we believe that the different measures may contain complementary information, and the combination of these measures may increase the AD classification performance compared to the separate measures. In this article, the combination of multiple anatomical MRI measures is utilized for the diagnosis of AD.

Kernel methods, such as support vector machines (SVM) [23], have been extensively used in many data analysis applications [24], [25]. Since single kernel methods might not be effective enough to handle diverse patterns and complex decision boundaries in real applications, multiple kernel learning (MKL) methods are proposed to achieve better capability and more flexibility in solving real-world challenges [26], [27], [28], [29], [30]. MKL can integrate different sources of feature spaces, thus providing a general framework for data fusion. It has found successful applications in genomic data fusion [31], [32], disease classification [33], [34], [35], etc. As for AD data fusion and classification, Zhang et al. [33] proposed a MKL method to generate mixed kernels by using MRI, PET and CSF features, and Liu et al. [35] proposed a MKL framework based on random Fourier features (RFF) and L_{21} norm to distinguish progressive MCI patients and HC and obtained a good classification performance.

In this article, we propose a novel AD diagnosis method by combining multiple measures through MKL. Our used anatomical measures include CGMV, CT, CSA, CC, CFI and SV. First, we construct six individual networks based on these six different anatomical measures and Automated Anatomical Labeling (AAL) atlas [36] for each subject. Then, for each individual network, we extract all node features and edge features, and denoted as node feature set and edge feature set, respectively. Therefore, we can obtain six node feature sets and six edge feature sets. Next, each feature within a feature set is ranked by F-score in descending order, and the top ranked features of each feature set are applied to MKBoost algorithm [30] to obtain the best classification accuracy. After obtaining the best classification accuracy, we can get the optimal feature subset and the corresponding classifier for each node or edge feature set. Afterwards, to investigate the classification performance with only node features, we propose a weighted multiple kernel learning (wMKL) framework to combine these six optimal node feature subsets, and obtain a combined classifier to achieve AD classification. Similarly, we can obtain the classification performance with only edge features. Finally, we combine both six optimal node feature subsets and six optimal edge feature subsets to further improve the classification performance.

2 Materials and Methods

2.1 Image Description and Preprocessing

A subset of the Alzheimers Disease Neuroimaging Initiative (ADNI) dataset [37] is used to evaluate our proposed method. This subset includes 710 subjects with T_1-weighted MRI images, which are composed of 200 patients with AD, 280 subjects with MCI (120 subjects that had MCI and converted to AD within 18 months (MCIc), 160 subjects with MCI that did not convert (MCInc)) and 230 HCs. The demographic information of 710 subjects is shown in Table 1, where MMSE stands for Mini Mental State Examination. For more details with the ADNI database, please see http://adni.loni.usc.edu/.

All T_1-weighted MRI images are preprocessed by using the Freesurfer image analysis suite (stable release version 5.3.0;

TABLE 1
Demographic Information of 710 Subjects from ADNI Database

Type	Number	Age	Gender (M/F)	MMSE
HC	230	77.13 ± 6.24	116/84	29.16 ± 0.82
MCInc	160	76.26 ± 5.35	89/71	27.56 ± 1.18
MCIc	120	75.95 ± 6.27	53/67	26.38 ± 1.52
MCI	280	76.11 ± 5.98	142/138	26.97 ± 1.34
AD	200	76.63 ± 5.91	122/108	23.54 ± 2.07

The values are denoted as mean $\pm$ standard deviation.

http://surfer.nmr.mgh.harvard.edu/). The image preprocessing procedures mainly include motion correction, intensity normalization, skull stripping and cerebellum removal.

2.2 Measures of Brain Regions

2.2.1 Measures of Cortical Regions

Several measures of cortical regions [16], [20], [21], [22], [38], [39], such as CGMV, CT, CSA, CC and CFI have been identified and used successfully to discriminate AD, MCI and HC. These measures have so far mainly been used separately. The full potential of these measures for AD diagnosis have thus not yet been used optimally. To further improve AD diagnosis, in this study, we integrate multiple measures of cortical regions including CGMV, CT, CSA, CC and CFI.

To calculate CGMV, CT, CSA, CC and CFI of cortical regions, anatomic reconstruction of the cortical surfaces of the T_1-weighted MRI images is performed using the Freesurfer software [40], [41]. Before anatomic reconstruction, all T_1-weighted MRI images are registered to AAL atlas as shown in Table 2. In the process of anatomic reconstruction, triangle meshes which represent the boundary of the white matter surface (the GM−WM interface) and the boundary of the pial surface (the GM−CSF interface) are generated by using deformation algorithms based on local intensity values [40] and geometrical and topological constraints [42]. Each triangular mesh consists of over 100,000 vertices for each hemisphere. The reconstruction procedure is repeated until accurate representations of white matter and pial surfaces were obtained. The reconstructed surfaces are used to calculate these measures. The volume between the white matter and pial surfaces is calculated as CGMV. The mean of the two shortest distances between the white matter and pial surfaces is calculated as CT. Winkler et al. [43] think that CSA of each vertex is equal to the mean of its surrounding triangles. Therefore, the sum of the areas of the triangles can be used to represent CSA of a ROI. The inverse of the length of the radius of the osculating circles in these directions is used as the curvature of a vertex in these directions [44]. Therefore, the mean of the curvature values in the two principal directions of the surface can be used to represent CC of a ROI. CFI, also called local gyrification index (lGI), is measured by surface-based 3D gyrification technique previously proposed by Schaer et al. [45]. The technique has been embedded in Freesurfer. The lGI at a given voxel on the cortical surface is computed as the ratio between the surface of a 25 mm radius circular ROI on the folded pial surface and the surface of the corresponding cortex's outer perimeter [45]. The amount of cortical folding (lGI) at each pial surface location reflects the amount of cortex buried within the sulcal folds in the surrounding area. The greater the value of the lGI, the more surfaces are buried in sulcal folds [46].

TABLE 2
Cortical and Subcortical Regions Defined in the Present Study

Cortical Regions (39 in each hemisphere)			
Regions	*Abbr.*	*Regions*	*Abbr.*
Precentral gyrus	PreCG	Superior frontal gyrus (dorsal)	SFGdor
Orbitofrontal cortex (superior)	ORBsup	Middle frontal gyrus	MFG
Orbitofrontal cortex (middle)	ORBmid	Inferior frontal gyrus, opercular part	IFGoper
Inferior frontal gyrus, triangular part	IFGtri	Orbitofrontal cortex (inferior)	ORBinf
Rolandic operculum	ROL	Supplementary motor area	SMA
Olfactory cortex	OLF	Superior frontal gyrus, medial	SFGmed
Orbitofrontal cortex (medial)	ORBmed	Gyrus rectus	REC
Insula	INS	Anterior cingulate gyrus	ACG
Median cingulate gyrus	MCG	Posterior cingulate gyrus	PCG
Parahippocampal gyrus	PHG	Calcarine fissure	CAL
Cuneus	CUN	Lingual gyrus	LING
Superior occipital gyrus	SOG	Middle occipital gyrus	MOG
Inferior occipital gyrus	IOG	Fusiform gyrus	FFG
Postcentral gyrus	PoCG	Superior parietal gyrus	SPG
Inferior parietal lobule	IPL	Supramarginal gyrus	SMG
Angular gyrus	ANG	Precuneus	PCUN
Paracentral lobule	PCL	Heschl gyrus	HES
Superior temporal gyrus	STG	Temporal pole: superior temporal gyrus	TPOsup
Middle temporal gyrus	MTG	Temporal pole: middle temporal gyrus	TPOmid
Inferior temporal gyrus	ITG		
Subcortical Regions (6 in each hemisphere)			
Regions	*Abbr.*	*Regions*	*Abbr.*
Lenticular nucleus, Putamen	PUT	Lenticular nucleus, Pallidum	PAL
Thalamus	THA	Caudate nucleus	CAU
Amygdala	AMYG	Hippocampus	HIP

2.2.2 Measure of Subcortical Regions

The SV has been used successfully to discriminate patients with AD from HCs [47]. In this study, we calculate the volumes of the subcortical structures using the FIRST in FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL) [48]. At the first stage, the nonlinear MNI-152 template is used to register all whole brain images. Furthermore, to achieve a more accurate registration, a subcortical mask (from AAL atlas as shown in Table 2) is used on the basis of the previous results. At the second stage, deformable meshes are used to model the shapes of the subcortical structures, and a structural segmentation method [49] is used to classify the boundary voxels as being part of the subcortical structure. At the third stage, the subcortical volumes are corrected by intracranial volume, which is obtained by FSL.

2.3 Construction of Individual Networks

A network is typically defined as $G = (V, E)$, where V is the set of vertices (or nodes) and E is the set of edges. In this article, we construct two kinds of individual networks: cortical network and subcortical network.

For cortical network, we can construct five kinds of individual networks based on the aforementioned five measures of cortical regions, denoted as G_{CGMV}, G_{CT}, G_{CSA}, G_{CC} and G_{CFI}. For G_{CGMV}, a node is a cortical region as defined by the AAL atlas and an edge is a certain similarity of the CGMV between a pair of nodes. In this study, the similarity of the CGMV between a pair of nodes is calculated as following:

$$s(a,b) = \frac{1}{d(a,b)+1}, \tag{1}$$

where $d(a,b)$ is defined as the difference between ROIs a and b

$$d(a,b) = |t(a) - t(b)|, \tag{2}$$

where $t(a)$ and $t(b)$ are the CGMV of ROIs a and b, respectively. The construction methods of other four individual cortical networks (G_{CT}, G_{CSA}, G_{CC} and G_{CFI}) are similar to that of G_{CGMV}. These five individual cortical networks share the same set of cortical regions and use the same method to calculate the similarity between a pair of nodes. However, they employ different definitions of $t(\cdot)$ that represents the CT, CSA, CC and CFI of a cortical region in G_{CT}, G_{CSA}, G_{CC} and G_{CFI}, respectively.

For the subcortical network, we only construct an individual network based on the aforementioned measure of subcortical regions, denoted as G_{SV}. The construction method of G_{SV} is similar to the aforementioned five individual networks. They use the same method to calculate the similarity between a pair of nodes. Their differences mainly include that the set of nodes in G_{SV} is all of subcortical regions instead of cortical regions and $t(\cdot)$ in G_{SV} represents the SV of a subcortical region.

It is worth mentioning that other functions could also be used to calculate the correlation between two ROIs. We choose the inverse proportional function because of its nonlinearity, monotonicity, boundness, and direct reflection to the dissimilarity. The smaller the difference between two ROIs, the larger their correlation is. If two ROIs have the same value, they have the strongest similarity whose correlation equals 1.

Based on the analysis above, we can obtain six individual networks for each subject. In this study, we will extract all nodes and edges of each individual network as the original features for AD classification. Based on these six individual networks, we can extract two kinds of feature sets: node feature set and edge feature set as follows:

- Node feature set: the sets of node features of G_{CGMV}, G_{CT}, G_{CSA}, G_{CC}, G_{CFI} and G_{SV} are denoted as N_{CGMV}, N_{CT}, N_{CSA}, N_{CC}, N_{CFI} and N_{SV}, respectively. The dimension of the features in N_{SV} is 12, and the dimension of the features in other five sets is 78.
- Edge feature set: the sets of edge features of G_{CGMV}, G_{CT}, G_{CSA}, G_{CC}, G_{CFI} and G_{SV} are denoted as E_{CGMV}, E_{CT}, E_{CSA}, E_{CC}, E_{CFI} and E_{SV}, respectively. The dimension of the features in E_{SV} is 66, and the dimension of the features in other five sets is 3,003.

2.4 Feature Ranking

Solving classification problems with data of high dimensionality is a challenging task due to the curse of dimensionality. This is particularly obvious for neuroimaging classification problems. With the presence of uninformative, irrelevant or redundant features, learning models tend to overfit and become less generalizable. Feature ranking is a useful and important means to identify relevant features for dimensionality reduction and improving generalization performance [50], [51]. In this study, we use the F-score [52] to rank all features within each feature set (such as N_{CGMV}, N_{CT}, E_{CC}, E_{CFI}, etc.).

F-score is a simple technique which measures the discrimination of two sets of real numbers. Given training vectors $\boldsymbol{x}_k$, $k = 1, \ldots, m$, if the number of positive instances and negative instances are n_+ and n_-, respectively, then the F-score of the ith feature is defined as

$$F(i) = \frac{\left(\overline{\boldsymbol{x}}_i^{(+)} - \overline{\boldsymbol{x}}_i\right)^2 + \left(\overline{\boldsymbol{x}}_i^{(-)} - \overline{\boldsymbol{x}}_i\right)^2}{\frac{1}{n_+-1}\sum_{k=1}^{n_+}\left(\boldsymbol{x}_{k,i}^{(+)} - \overline{\boldsymbol{x}}_i^{(+)}\right)^2 + \frac{1}{n_--1}\sum_{k=1}^{n_-}\left(\boldsymbol{x}_{k,i}^{(-)} - \overline{\boldsymbol{x}}_i^{(-)}\right)^2}, \tag{3}$$

where $\overline{\boldsymbol{x}}_i$, $\overline{\boldsymbol{x}}_i^{(+)}$, $\overline{\boldsymbol{x}}_i^{(-)}$ are the average of the ith feature of the whole, positive, and negative data sets, respectively; $\boldsymbol{x}_{k,i}^{(+)}$ is the ith feature of the kth positive instance, and $\boldsymbol{x}_{k,i}^{(-)}$ is the ith feature of the kth negative instance. The numerator indicates the discrimination between the positive and negative sets, and the denominator indicates the one within each of the two sets. The larger the F-score is, the more likely this feature is more discriminative.

2.5 Classification with Multiple Kernel Learning

Multiple Kernel Learning is a promising family of machine learning algorithms with multiple kernel functions for various challenging data mining tasks [26], [27], [30]. In this study, we propose a two-step MKL method to achieve the final classification.

First, we use MKBoost-S2 algorithm proposed by Hao et al. [30] to obtain the optimal feature subset and the optimal classifier for each feature set defined in Section 2.3. Actually with a specific feature set and a number of base kernels, MKBoost-S2 algorithm can generate a hybrid kernel

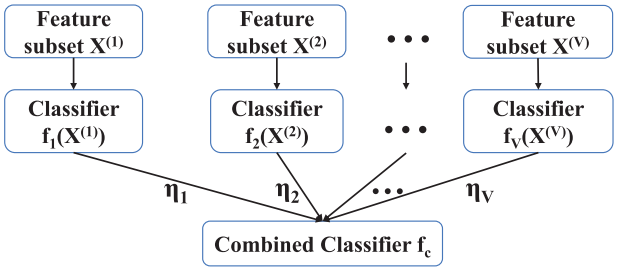

Fig. 1. Schematic illustration of the proposed MKL framework.

whose corresponding classifier is with the best accuracy. For more details about MKBoost-S2 algorithm, please see [30]. Since the polynomial kernel has good global performance and the Gaussian kernel has good local performance, in this study, we define a set of 13 base kernels for MKBoost-S2 algorithm (i.e., M = 13) as follows:

- The polynomial kernels are defined as

$$k_{polynomial}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i{}^T \boldsymbol{x}_j + 1)^d, \tag{4}$$

  where $\boldsymbol{x}_i{}^T$ is the transpose of $\boldsymbol{x}_i$, and d is the degree of the polynomial kernel. In this study, we use polynomial kernels with 3 different values of d (i.e., $d = \{1, 2, 3\}$) for each feature set.
- The Gaussian kernels are defined as

$$k_{Gaussion}(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\sigma^2}), \tag{5}$$

  where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are two feature vectors, and σ is the width of the Gaussian kernel. In this study, we use Gaussian kernels with 10 different values of σ (i.e., $\sigma = \left\{2^{-4}, 2^{-3}, \ldots, 2^4, 2^5\right\}$) for each feature set.

For specific feature set N_{CGMV}, we construct K different feature subsets with its top p_i $(i = 1, \ldots, k)$ features ranked in Section 2.4, where p_i can be 5, 10, 15, and so on. With all 13 defined based kernels, MKBoost-S2 algorithm generates a classifier for each of these K feature subsets, As a result, K classifiers are obtained. Among K classifiers, the one with the best accuracy is the optimal classifier (denoted as KN_{CGMV}) and its corresponding feature subset is the optimal feature subset (denoted as FN_{CGMV}). Similarly, for other feature sets, i.e., N_{CT}, N_{CSA}, N_{CC}, N_{CFI}, N_{SV}, E_{CGMV}, E_{CT}, E_{CSA}, E_{CC}, E_{CFI} and E_{SV}, we can also get their corresponding optimal feature subsets, i.e., FN_{CT}, FN_{CSA}, FN_{CC}, FN_{CFI}, FN_{SV}, FE_{CGMV}, FE_{CT}, FE_{CSA}, FE_{CC}, FE_{CFI} and FE_{SV}, and their corresponding optimal classifiers, i.e., KN_{CT}, KN_{CSA}, KN_{CC}, KN_{CFI}, KN_{SV}, KE_{CGMV}, KE_{CT}, KE_{CSA}, KE_{CC}, KE_{CFI} and KE_{SV}, respectively.

Second, a weighted MKL (wMKL) framework is proposed to construct a combined classifier by using the optimal feature subsets and the corresponding optimal classifiers to investigate classification performance as shown in Fig. 1.

Suppose there are V feature subsets, $X = \{X^{(1)}, \ldots, X^{(V)}\}$, corresponding to V classifiers, $\{f_1(X^{(1)}), \ldots, f_V(X^{(V)})\}$, the combined classifier (denoted as f_c) can be computed by

$$f_c = sign\left(\sum_{v=1}^{V} \eta_v f_v(X^{(v)})\right), \tag{6}$$

where η_v represents the weight of classifier v, which can be either assigned prior to learning or determined in the learning procedure. In this study, a three-step wMKL method is proposed to achieve the final classification. We first consider six optimal node feature subsets and the corresponding six optimal classifiers. In this case, We can generate a combined node classifier KN by integrating KN_{CGMV}, KN_{CT}, KN_{CSA}, KN_{CC}, KN_{CFI} and KN_{SV} with their accuracies as the weights via Equation (6). Then, by using a similar process for getting KN, we can get a combined edge classifier KE by integrating KE_{CGMV}, KE_{CT}, KE_{CSA}, KE_{CC}, KE_{CFI} and KE_{SV}. Finally, we use the accuracies of KN and KE as the weights to get a combined final classifier KNE via Equation (6) for all node and edge feature sets to perform the final classification.

In this study, support vector machine (SVM) [53], [54] is adopted to perform classifications with different kernels and implemented by the LIBSVM library [55], with a default value for the parameter C (= 1). All classification procedures are performed via a 10-fold cross-validation (CV) and repeated 100 times. The source code of our method can be downloaded at https://github.com/bioinfomaticsCSU/AD_Classification.

2.6 Evaluation

An overview of the proposed classification framework is depicted in Fig. 2.

To investigate the performance of our proposed classification framework, four classification types in AD diagnosis are conducted as follows:

- T1: AD/HC classification
- T2: AD/MCI classification
- T3: MCI/HC classification
- T4: MCIc/MCInc classification

Accuracy (ACC), sensitivity (SEN), and specificity (SPE) are used to evaluate the classification performance of our proposed method. These three metrics are calculated by the following formulas:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$SEN = \frac{TP}{TP + FN} \tag{8}$$

$$SPE = \frac{TN}{TN + FP}, \tag{9}$$

where TP, FP, TN, and FN are the number of true positives, false positives, true negatives, and false negatives, respectively. In addition, the area under receiver operating characteristic (ROC) curve (AUC) is also reported to evaluate the overall performance of a classification method. The higher the AUC value, the better the classification performance of the classification method is.

3 Results and Discussion

3.1 The Classification Accuracy of Each Feature Set

To obtain the best classification accuracy of each feature set for T1, T2, T3 and T4, we first investigate the classification accuracy of each feature set. All features within each feature set are ranked according to F-score in descending order. The

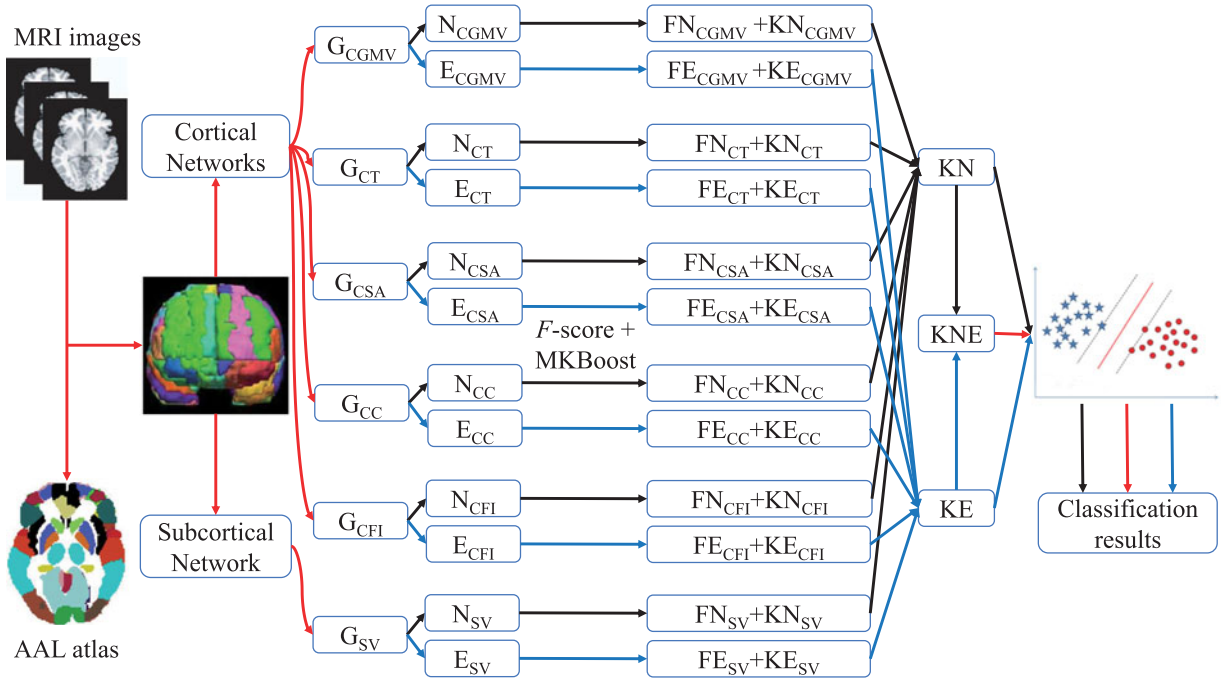

Fig. 2. An overall flowchart for AD classification by combining multiple anatomical MRI measures through MKL. First, we construct five cortical networks (i.e., G_{CGMV}, G_{CT}, G_{CSA}, G_{CC}, and G_{CFI}) and one subcortical network (i.e., G_{SV}) by using different anatomical measures and AAL atlas for each subject. Then, we extract node feature sets (N_{CGMV}, N_{CT}, N_{CSA}, N_{CC}, N_{CFI}, and N_{SV}) and edge feature sets (E_{CGMV}, E_{CT}, E_{CSA}, E_{CC}, E_{CFI}, and E_{SV}) from these six individual networks. Next, all features within each of 12 feature sets are ranked by F-score in descending order, and the top ranked features of each feature set are applied to MKBoost algorithm to obtain the best classification accuracy. After obtaining the best classification accuracy, we can get the optimal feature subset (i.e., FN_{CT}, FN_{CSA}, FN_{CC}, FN_{CFI}, FN_{SV}, FE_{CGMV}, FE_{CT}, FE_{CSA}, FE_{CC}, FE_{CFI}, and FE_{SV}) and the corresponding classifier (i.e., KN_{CGMV}, KN_{CT}, KN_{CSA}, KN_{CC}, KN_{CFI}, KN_{SV}, KE_{CGMV}, KE_{CT}, KE_{CSA}, KE_{CC}, KE_{CFI}, and KE_{SV}) for each feature subset. Afterwards, to investigate the classification performance with only node features, we propose a weighted MKL framework to combine these six optimal node feature subsets, and obtain a combined classifier (KN) to achieve AD classification. Similarly, we can obtain the classification performance with only edge features by using a combined classifier (KE). Finally, we combine both six optimal node feature subsets and six optimal edge feature subsets to generate a combined classifier (KNE) to further improve the classification performance.

top ranked features of each feature set are utilized for the classification features of MKBoost-S2 algorithm. As shown in Figs. 3 and 4, the mean classification accuracy of each feature set is obtained, and Fig. 3 shows the mean classification accuracy of each node feature set while Fig. 4 shows the mean classification accuracy of each edge feature set.

From Fig. 3, we can determine not only the best classification accuracy of each node feature set, but also the optimal node feature subset and the corresponding classifier for each node feature set. Similarly, from Fig. 4, we also can determine the best classification accuracy, the optimal edge feature subset and the corresponding classifier for each edge feature set. Table 3 shows the best classification accuracy and the number of features in the optimal feature subset of the corresponding classifier.

3.2 Classification Performance of Node Feature Set and Edge Feature Set

First, we only focus on the classification performance of node feature sets (Black line in Fig. 2). Based on the best classification accuracy and the corresponding classifier of each optimal node feature set, we use the aforementioned weighted MKL framework to combine these six classifiers to generate a combined classifier (KN). Therefore, we can obtain four specific combined classifiers for T1, T2, T3 and T4. Four SVMs with corresponding specific combined classifiers are utilized for T1, T2, T3 and T4, respectively. The results of T1, T2, T3 and T4 based on combined node feature sets are shown in Table 4. By the comparison of Tables 3 and 4, we can easily find that the classification accuracy of combined node feature sets is superior to any optimal node feature set for T1, T2, T3 and T4.

Second, we only focus on the classification performance of edge feature sets (Blue line in Fig. 2). Similar to the procedure of node feature sets, we also can obtain a specific combined classifier (KE) for each classification type (i.e., T1, T2, T3 and T4). Similarly, Four SVMs with corresponding specific combined classifiers are utilized for T1, T2, T3 and T4, respectively. The results of T1, T2, T3 and T4 based on combined edge feature sets are shown in Table 4. By the comparison of Tables 3 and 4, we also can see that the classification accuracy of combined edge feature sets is superior to any optimal edge feature set for T1, T2, T3 and T4.

As can be seen from Table 4, for each classification type, the classification performance of combined edge feature sets is better than that of corresponding combined node feature sets. The results indicate that the correlation between brain regions are more significant than that of brain regions themselves in AD classification. To further improve the classification performance, we consider not only combined node feature sets, but also combined edge feature sets. By using the proposed weighted MKL framework, for each classification type, we can generate a combined classifier (KNE) of KN and KE, and use the KNE to perform the final classification (Red line in Fig. 2). The classification results from the corresponding KNE of each classification

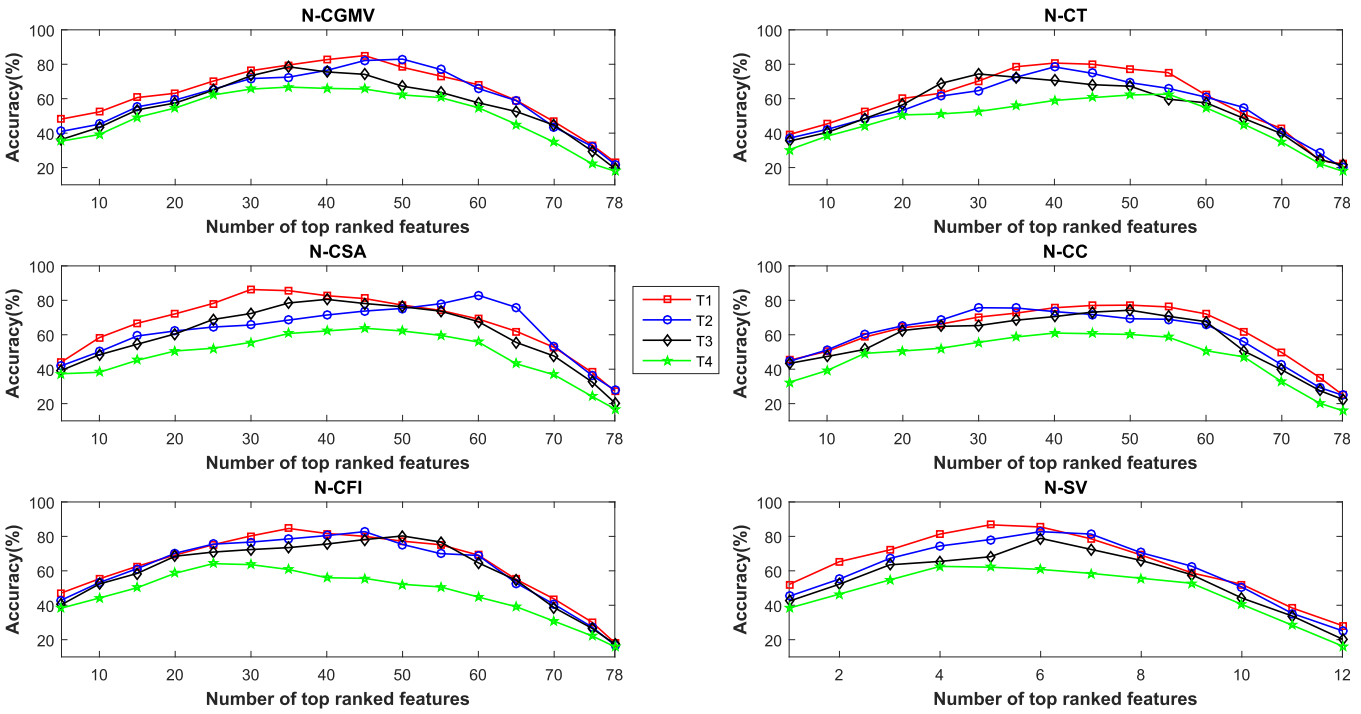

Fig. 3. The classification accuracy of each node feature set for T1, T2, T3, and T4. All features within each node feature set is ranked according to F-score in descending order. N-CGMV, N-CT, N-CSA, N-CC, N-CFI, and N-SV are N_{CGMV}, N_{CT}, N_{CSA}, N_{CC}, N_{CFI}, and N_{SV}, respectively.

type are shown in Table 4. In Table 4, we can easily see that for each classification type, the classifier KNE is better than both corresponding classifiers: KN and KE for all classifications T1, T2, T3, and T4.

Based on the above analysis, the classification performance of multiple measures is better than single measure, which indicates that the different measures provide complementary information for AD diagnosis.

3.3 Comparison with Existing Classification Methods

Recently, some existing methods have achieved good results for AD classification. For example, Chupin et al. [13] proposed to use the hippocampal volumes to automatically discriminate between populations of AD, MCI and HC; Ahmed et al. [14] proposed an automatic classification framework based on the fusion of both hippocampus visual

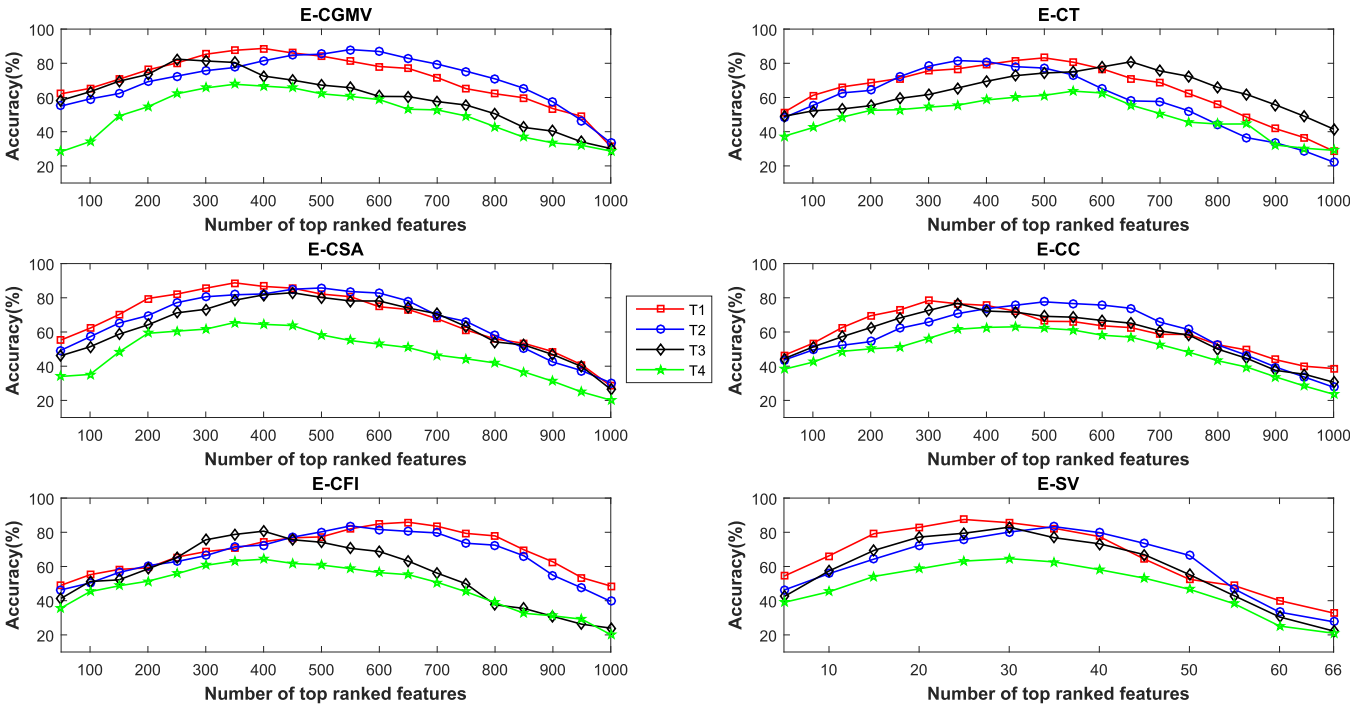

Fig. 4. The classification accuracy of each edge feature set for T1, T2, T3, and T4. All features within each edge feature set is ranked according to F-score in descending order. E-CGMV, E-CT, E-CSA, E-CC, E-CFI, and E-SV are E_{CGMV}, E_{CT}, E_{CSA}, E_{CC}, E_{CFI}, and E_{SV}, respectively.

TABLE 3
The Best Classification Accuracy and the Number of Features in the Optimal Feature Subset of the Corresponding Classifier

Type	Node feature set			Edge feature set		
	Classifier	ACC(%)	Features	Classifier	ACC(%)	Features
T1	KN_{CGMV}	85.02	45	KE_{CGMV}	88.69	400
	KN_{CT}	80.69	40	KE_{CT}	83.17	500
	KN_{CSA}	86.28	30	KE_{CSA}	88.74	350
	KN_{CC}	77.21	50	KE_{CC}	78.35	300
	KN_{CFI}	84.59	35	KE_{CFI}	85.89	650
	KN_{SA}	86.73	5	KE_{SA}	87.48	25
T2	KN_{CGMV}	82.98	50	KE_{CGMV}	88.01	550
	KN_{CT}	78.46	40	KE_{CT}	81.59	350
	KN_{CSA}	82.91	60	KE_{CSA}	85.76	500
	KN_{CC}	75.53	30	KE_{CC}	77.62	500
	KN_{CFI}	82.76	45	KE_{CFI}	83.73	550
	KN_{SA}	82.75	6	KE_{SA}	83.27	35
T3	KN_{CGMV}	78.49	35	KE_{CGMV}	82.42	250
	KN_{CT}	74.35	30	KE_{CT}	80.89	650
	KN_{CSA}	80.55	40	KE_{CSA}	82.96	450
	KN_{CC}	74.25	50	KE_{CC}	76.76	350
	KN_{CFI}	80.25	50	KE_{CFI}	80.65	400
	KN_{SA}	78.86	6	KE_{SA}	82.97	30
T4	KN_{CGMV}	66.74	35	KE_{CGMV}	67.74	350
	KN_{CT}	62.65	55	KE_{CT}	63.73	550
	KN_{CSA}	63.78	45	KE_{CSA}	65.53	350
	KN_{CC}	60.98	40	KE_{CC}	63.02	450
	KN_{CFI}	64.16	25	KE_{CFI}	64.25	400
	KN_{SA}	62.58	4	KE_{SA}	64.59	30

TABLE 4
The Classification Performance for Different Feature Sets

Type	Classifier	ACC(%)	SEN(%)	SPE(%)	AUC
T1	KN	92.68	91.82	93.75	0.9428
	KE	93.57	92.03	94.16	0.9639
	KNE	95.24	94.26	95.74	0.9754
T2	KN	89.55	90.97	88.24	0.9246
	KE	90.74	91.16	88.98	0.9287
	KNE	90.85	91.77	89.56	0.9355
T3	KN	85.01	88.24	80.19	0.8852
	KE	85.97	86.49	84.77	0.8916
	KNE	86.35	89.49	85.68	0.9107
T4	KN	70.06	66.59	72.15	0.7249
	KE	72.97	68.33	73.54	0.7432
	KNE	74.28	71.51	76.46	0.7885

TABLE 5
Comparison to Existing Works Using MRI Data of ADNI for AD/HC Classification

Method	ACC(%)	SEN(%)	SPE(%)	AUC
Chupin et al., 2009 [13]	80.51	78.76	82.16	0.7851
Ahmed et al., 2015 [14]	86.40	77.61	93.28	0.8487
Suk et al., 2014 [56]	93.05	90.86	94.57	0.9475
Khedher et al., 2015 [58]	88.96	92.35	86.24	0.9256
Dai et al., 2013 [57]	90.81	92.59	90.33	0.9429
Proposed method	95.24	94.26	95.74	0.9754

TABLE 6
Comparison to Existing Works Using MRI Data of ADNI for AD/MCI Classification

Method	ACC(%)	SEN(%)	SPE(%)	AUC
Chupin et al., 2009 [13]	73.48	70.35	76.98	0.7328
Ahmed et al., 2015 [14]	74.51	77.94	71.23	0.7562
Suk et al., 2014 [56]	88.98	82.11	90.65	0.9007
Khedher et al., 2015 [58]	84.59	88.75	83.07	0.8859
Dai et al., 2013 [57]	85.92	82.46	87.59	0.8743
Proposed method	90.85	91.77	89.56	0.9355

TABLE 7
Comparison to Existing Works Using MRI Data of ADNI for MCI/HC Classification

Method	ACC(%)	SEN(%)	SPE(%)	AUC
Chupin et al., 2009 [13]	71.94	65.79	74.18	0.7155
Ahmed et al., 2015 [14]	76.29	72.30	81.53	0.7677
Suk et al., 2014 [56]	83.67	96.79	57.28	0.8203
Khedher et al., 2015 [58]	82.41	84.12	80.48	0.8134
Dai et al., 2013 [57]	81.92	78.51	88.34	0.8118
Proposed method	86.35	89.49	85.68	0.9107

TABLE 8
Comparison to Existing Works Using MRI Data of ADNI for MCIc/MCInc Classification

Method	ACC(%)	SEN(%)	SPE(%)	AUC
Chupin et al., 2009 [13]	64.21	69.74	62.45	0.6638
Ahmed et al., 2015 [14]	68.72	67.38	70.69	0.6814
Suk et al., 2014 [56]	72.86	40.55	88.49	0.7123
Khedher et al., 2015 [58]	70.11	68.61	74.16	0.7076
Dai et al., 2013 [57]	71.04	65.98	75.56	0.7086
Proposed method	74.28	71.51	76.46	0.7885

features and cerebrospinal fluid volume for AD recognition in structural MRI; Suk et al. [56] used Deep Boltzmann Machine, a deep network with a restricted Boltzmann machine as a building block, to find a latent hierarchical feature representation from a 3D patch for AD/MCI diagnosis; Dai et al. [57] presented a kernel-based method to establish an individual network for each subject based on mean cortical thickness and used the network edge features to make prediction of AD/NC through the sophisticated machine learning technology; Khedher et al. [58] combined the different brain tissues to improve the classification performance of AD.

To demonstrate the superiority of the proposed method, we compare our results with the above mentioned five methods on AD classification. For fair comparisons, the data sets of these experiments are the same. Although Suk et al. [56] proposed the multi-modality based approach, only the results based on MRI data are reported in this study. Tables 5, 6, 7 and 8 present the comparison results for T1, T2, T3 and T4, respectively. As can be seen from Tables 5, 6, 7, and 8, the methods based on whole brain substantially outperform both single ROI method proposed by Chupin et al. [13] and multiple ROIs method proposed by Ahmed et al. [14], in terms of ACC, SEN, SPE and AUC. Although the sensitivity (SEN) of Suk et al. [56] method is low (40.55 percent), its other three metrics (ACC, SPE and AUC) are relatively high. These results indicate that AD may be related to many regions of the brain, rather than the

TABLE 9
The Best AUC Value of Different Number of Measures for Different Classification

Type	$N_d = 2$	$N_d = 3$	$N_d = 4$	$N_d = 5$	$N_d = 6$
T1	0.9594 ③⑥	0.9788 ①③⑥	0.9651 ①③④⑥	0.9731 ①②③⑤⑥	0.9754
T2	0.9287 ①③	0.9322 ①⑤⑥	0.9401 ①②③④	0.9376 ①③④⑤⑥	0.9355
T3	0.9034 ③⑤	0.9016 ③⑤⑥	0.8942 ①②③⑤	0.9009 ①②④⑤⑥	0.9107
T4	0.7512 ①⑤	0.7829 ①⑤⑥	0.7649 ①②③⑤	0.7776 ①②③⑤⑥	0.7885

①↦*CGMV;* ②↦*CT;* ③↦*CSA;* ④↦*CC;* ⑤↦*CFI;* ⑥↦*SV*

specific one or several regions, which is in agreement with many existing reports [56], [57], [58]. Meanwhile, the classification performance of our proposed method is also superior to Suk et al. [56], Dai et al. [57] method and Khedher et al. [58] method, which are whole brain methods.

The AUC value has been widely used to evaluate the overall performance of classification method. The greater the AUC value, the better the classification performance of the proposed method is. As shown in Tables 5, 6, 7 and 8, the AUC values of our proposed method are superior to the other five existing methods for the above four classifications. The results demonstrate that our proposed method is efficient and strongly robust.

3.4 Comparison with Different Number of Measures

To answer the question: how many of brain measures (such as CGMV, CT, CSA, etc.) are needed for a specific classification, we have done a series of experiments based on different number of brain measures (i.e., $N_d = \{2, 3, 4, 5, 6\}$), and have also obtained the best AUC value of the combination of different N_d for different classification as shown in Table 9. As can be seen from Table 9, T1 gets the highest AUC value (0.9788) with CGMV, CSA and SV measures ($N_d = 3$) among $N_d = \{2, 3, 4, 5, 6\}$, and T2 gets the highest AUC value (0.9401) with CGMV, CT, CSA and CC measures ($N_d = 4$) among $N_d = \{2, 3, 4, 5, 6\}$; however, T3 gets the highest AUC value (0.9107) with all six measures ($N_d = 6$) among $N_d = \{2, 3, 4, 5, 6\}$, and T4 also gets the highest AUC value (0.7885) with all six measures ($N_d = 6$) among $N_d = \{2, 3, 4, 5, 6\}$.

In summary, using all six measures gets a relatively good classification performance for T1, T2, T3 and T4 as shown in Table 9. However, for some classification, using fewer measures can get better classification performance. For example, T1 gets the highest AUC value with CGMV, CSA and SV measures, and T2 gets the highest AUC value with CGMV, CT, CSA and CC measures as shown in Table 9.

4 Conclusion

In this study, we have developed an Alzheimer's disease diagnosis method by combining multiple anatomical MRI measures through multiple kernel learning. Based on AAL atlas and six different anatomical measures, we first construct six individual networks for each subject and extract two kinds of feature sets: node feature set and edge feature set. Then, we use a F-score method to rank features of each feature set. Afterwards, the top ranked features are applied to MKBoost-S2 algorithm to obtain the best classification accuracy, the corresponding classifier and the optimal feature subset. Finally, a weighted MKL framework is proposed to combine node feature set, edge feature set, and both node and edge feature sets, respectively to perform the final classification. Experimental results on the MRI data from ADNI database have showed that our proposed method is efficient and can achieve the better classification performance. Our results further demonstrate that clinical AD diagnosis could benefit from calculating multiple measures from an anatomical MRI scan and incorporate these all in an automated analysis.

Acknowledgments

The authors would like to express their gratitude for the support from the National Natural Science Foundation of China under Grant No.61232001, No.61420106009 and No.61622213, and Doctoral Student Independent Exploration Innovative Project No.2015zzts051.

References

[1] A. Association, "2014 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 10, no. 2, pp. e47–e92, 2014.
[2] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *J. Internal Med.*, vol. 256, no. 3, pp. 183–194, 2004.
[3] J. Nettiksimmons, C. DeCarli, S. Landau, L. Beckett, and A. D. N. Initiative, "Biological heterogeneity in ADNI amnestic mild cognitive impairment," *Alzheimer's Dementia*, vol. 10, no. 5, pp. 511–521, 2014.
[4] M. Bond, et al., "The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of alzheimers disease (review of technology appraisal no. 111): A systematic review and economic model," *Health Technol. Assessment*, vol. 16, no. 21, pp. 1–470, 2012.
[5] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient multilevel brain tumor segmentation with integrated Bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 629–640, May 2008.
[6] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of MRI-based brain tumor segmentation methods," *Tsinghua Sci. Technol.*, vol. 19, no. 6, pp. 578–595, 2014.
[7] M. Nieuwenhuis, N. E. van Haren, H. E. H. Pol, W. Cahn, R. S. Kahn, and H. G. Schnack, "Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples," *Neuroimage*, vol. 61, no. 3, pp. 606–612, 2012.
[8] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, and J. Wang, "Applications of deep learning to MRI images: A survey," *Big Data Mining Anal.*, vol. 1, no. 1, 2017.
[9] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimers disease using whole brain hierarchical network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2016. [Online]. Available: https://doi.org/10.1109/TCBB.2016.2635144
[10] J. Liu, J. Wang, H. Bin, F.-X. Wu, and Y. Pan, "Alzheimers disease classification based on individual hierarchical networks constructed with 3D texture features," *IEEE Trans. Nanobiosci.*, 2017. [Online]. Available: https://doi.org/10.1109/TNB.2017.2707139
[11] A. G. Filler, "The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI," *Internet J. Neurosurgery*, vol. 7, no. 1, pp. 5–35, 2010.
[12] P. Visser, F. Verhey, P. Hofman, P. Scheltens, and J. Jolles, "Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment," *J. Neurology Neurosurgery Psychiatry*, vol. 72, no. 4, pp. 491–497, 2002.
[13] M. Chupin, et al., "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, no. 6, pp. 579–587, 2009.

[14] O. B. Ahmed, et al., "Classification of Alzheimers disease subjects from MRI using hippocampal visual features," *Multimedia Tools Appl.*, vol. 74, no. 4, pp. 1249–1266, 2015.
[15] K. A. Linn, B. Gaonkar, T. D. Satterthwaite, J. Doshi, C. Davatzikos, and R. T. Shinohara, "Control-group feature normalization for multivariate pattern analysis of structural MRI data using the support vector machine," *NeuroImage*, vol. 132, pp. 157–166, 2016.
[16] G. Karas, et al., "Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease," *Neuroimage*, vol. 23, no. 2, pp. 708–716, 2004.
[17] V. Singh, H. Chertkow, J. P. Lerch, A. C. Evans, A. E. Dorr, and N. J. Kabani, "Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease," *Brain*, vol. 129, no. 11, pp. 2885–2893, 2006.
[18] M. Bobinski, et al., "MRI of entorhinal cortex in mild Alzheimer's disease," *Lancet*, vol. 353, no. 9146, pp. 38–40, 1999.
[19] J. H. Morra, et al., "Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls," *Neuroimage*, vol. 43, no. 1, pp. 59–68, 2008.
[20] D. M. Cash, et al., "Cortical folding analysis on patients with Alzheimers disease and mild cognitive impairment," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2012, pp. 289–296.
[21] E. Westman, C. Aguilar, J.-S. Muehlboeck, and A. Simmons, "Regional magnetic resonance imaging measures for multivariate analysis in Alzheimers disease and mild cognitive impairment," *Brain Topography*, vol. 26, no. 1, pp. 9–23, 2013.
[22] C. G. Schwarz, et al., "A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity," *NeuroImage: Clinical*, vol. 11, pp. 802–812, 2016.
[23] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[24] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[25] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[26] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, vol. 7, pp. 1531–1565, 2006.
[27] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, 2008.
[28] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, 2011.
[29] M. Gonen and S. Kaski, "Kernelized Bayesian matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2047–2060, Oct. 2014.
[30] H. Xia and S. C. Hoi, "MKBoost: A framework of multiple kernel boosting," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1574–1586, Jul. 2013.
[31] W. Lan, J. Wang, M. Li, J. Liu, F.-X. Wu, and Y. Pan, "Predicting microRNA-disease associations based on improved microrna and disease similarities," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2016. [Online]. Available: https://doi.org/10.1109/TCBB.2016.2586190
[32] W. Lan, et al., "LDAP: A web server for lncRNA-disease association prediction," *Bioinf.*, vol. 33, no. 3, pp. 458–460, 2016.
[33] D. Zhang, et al., "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.
[34] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Frontiers Syst. Neuroscience*, vol. 6, 2012, Art. no. 63.
[35] F. Liu, L. Zhou, C. Shen, and J. Yin, "Multiple kernel learning in the primal for multimodal Alzheimers disease classification," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 984–990, May 2014.
[36] N. Tzourio-Mazoyer, et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
[37] M. C. Carrillo, L. J. Bain, G. B. Frisoni, and M. W. Weiner, "Worldwide Alzheimer's disease neuroimaging initiative," *Alzheimer's Dementia*, vol. 8, no. 4, pp. 337–342, 2012.
[38] O. Querbes, et al., "Early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve," *Brain*, vol. 132, no. 8, pp. 2036–2047, 2009.
[39] B. C. Dickerson, et al., "Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area," *Neurobiology Aging*, vol. 30, no. 3, pp. 432–440, 2009.
[40] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis: I. segmentation and surface reconstruction," *Neuroimage*, vol. 9, no. 2, pp. 179–194, 1999.
[41] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system," *Neuroimage*, vol. 9, no. 2, pp. 195–207, 1999.
[42] B. Fischl, A. Liu, and A. M. Dale, "Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 70–80, Jan. 2001.
[43] A. M. Winkler, et al., "Measuring and comparing brain cortical surface area and other areal quantities," *Neuroimage*, vol. 61, no. 4, pp. 1428–1443, 2012.
[44] L. Ronan, et al., "Intrinsic curvature: A marker of millimeter-scale tangential cortico-cortical connectivity?" *Int. J. Neural Syst.*, vol. 21, no. 05, pp. 351–366, 2011.
[45] M. Schaer, M. B. Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J.-P. Thiran, "A surface-based approach to quantify local cortical gyrification," *IEEE Trans. Med. Imag.*, vol. 27, no. 2, pp. 161–170, Feb. 2008.
[46] M. Schaer, M. B. Cuadra, N. Schmansky, B. Fischl, J.-P. Thiran, and S. Eliez, "How to measure cortical folding from MR images: A step-by-step tutorial to compute local gyrification index," *J. Visualized Experiments*, vol. 59, pp. e3417–e3417, 2012.
[47] P. P. D. M. Oliveira Jr., R. Nitrini, G. Busatto, C. Buchpiguel, J. R. Sato, and E. Amaro Jr., "Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease," *J. Alzheimer's Disease*, vol. 19, no. 4, pp. 1263–1272, 2010.
[48] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, vol. 56, no. 3, pp. 907–922, 2011.
[49] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[50] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
[51] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
[52] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction*. Berlin, Germany: Springer, 2006, pp. 315–324.
[53] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[54] I. Steinwart and A. Christmann, *Support Vector Machines*. Berlin, Germany: Springer, 2008.
[55] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
[56] H.-I. Suk, S.-W. Lee, D. Shen, and A. D. N. Initiative, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.
[57] D. Dai, H. He, J. T. Vogelstein, and Z. Hou, "Accurate prediction of AD patients using cortical thickness networks," *Mach. Vis. Appl.*, vol. 24, no. 7, pp. 1445–1457, 2013.
[58] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, F. Segovia, and A. Disease Neuroimaging Initiative, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomput.*, vol. 151, pp. 139–150, 2015.

Jin Liu is currently working toward the PhD degree in the School of Information Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis, machine learning, and pattern recognition.

Jianxin Wang (M'01-SM'14) received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the chair of and a professor in the Department of Computer Science, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network. He is a senior member of the IEEE.

Zhenjun Tang received the BS and MEng degrees from Guangxi Normal University, Guilin, P.R. China, in 2003 and 2006, respectively, and the PhD degree from Shanghai University, Shanghai, P.R. China, in 2010. He is now a professor in the Department of Computer Science, Guangxi Normal University. His research interests include image processing and multimedia security. He has contributed more than 40 international journal papers. He holds six China patents. He is a senior member of the China Computer Federation (CCF) and also a reviewer of more than 20 SCI-indexed journals, such as the IEEE journals, IET journals, Elsevier journals, Springer journals, and Taylor & Francis journals.

Bin Hu received the MSc degree in computer science from the Beijing University of Technology and the PhD degree in computer science from Institute of Computing Technology, Chinese Academy of Science. He is a professor and dean of the School of Information Science and Engineering, Lanzhou University, the leader of the Intelligent Contextual Computing Group in the Pervasive Computing Centre, Reader, Birmingham City University, visiting professor with the Beijing University of Posts and Telecommunications, China, and at ETH Switzerland. His research interests include pervasive computing, psycho-physiological computing, cooperative work, and Semantic Web.

Fang-Xiang Wu (M'06-SM'11) received the BSc and MSc degree in applied mathematics, both from the Dalian University of Technology, Dalian, China, in 1990 and 1993, respectively, and the first PhD degree in control theory and its applications from Northwestern Polytechnical University, Xi'an, China, in 1998, and the second PhD degree in biomedical engineering from the University of Saskatchewan (U of S), Saskatoon, Canada, in 2004. From 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, 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. He is a senior member of the IEEE.

Yi Pan received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, Pennsylvania, in 1991. He is a regents' professor of the computer science and an interim associate dean and chair of the biology with Georgia State University, Georgia. 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 the Computer Science Department from 2005-2013. He is also a visiting Changjiang Chair Professor with Central South University, China. His profile has been featured as a distinguished alumnus in both *Tsinghua Alumni Newsletter* and the *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 more than 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 8,800 times. He has served as an editor-in-chief or editorial board member for 15 journals including seven IEEE Transactions. He is the recipient of many awards including IEEE Transactions Best Paper Award, four other international conference or journal Best Paper Awards, four IBM Faculty Awards, two 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 more than 50 international conferences around the world.