Hierarchical Fusion of Features and Classifier Decisions for Alzheimer's Disease Diagnosis

Manhua Liu[1,2], **Daoqiang Zhang**[2,3], **Dinggang Shen**[2,*], and **the Alzheimer's Disease Neuroimaging Initiative**

[1]Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, China

[2]Department of Radiology and BRIC, IDEA Lab, University of North Carolina at Chapel Hill, North Carolina

[3]Department of Computer Science and Engineering, Nanjing University of Aeronautics and, Astronautics, China

Abstract

Pattern classification methods have been widely investigated for analysis of brain images to assist the diagnosis of Alzheimer's disease (AD) and its early stage such as mild cognitive impairment (MCI). By considering the nature of pathological changes, a large number of features related to both local brain regions and interbrain regions can be extracted for classification. However, it is challenging to design a single global classifier to integrate all these features for effective classification, due to the issue of small sample size. To this end, we propose a hierarchical ensemble classification method to combine multilevel classifiers by gradually integrating a large number of features from both local brain regions and interbrain regions. Thus, the large-scale classification problem can be divided into a set of small-scale and easier-to-solve problems in a bottom-up and local-to-global fashion, for more accurate classification. To demonstrate its performance, we use the spatially normalized grey matter (GM) of each MR brain image as imaging features. Specifically, we first partition the whole brain image into a number of local brain regions and, for each brain region, we build two low-level classifiers to transform local imaging features and the inter-region correlations into high-level features. Then, we generate multiple high-level classifiers, with each evaluating the high-level features from the respective brain regions. Finally, we combine the outputs of all high-level classifiers for making a final classification. Our method has been evaluated using the baseline MR images of 652 subjects (including 198 AD patients, 225 MCI patients, and 229 normal controls (NC)) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The experimental results show that our classification method can achieve the accuracies of 92.0% and 85.3% for classifications of AD versus NC and MCI versus NC, respectively, demonstrating very promising classification performance compared to the state-of-the-art classification methods.

*Correspondence to: Dinggang Shen, Department of Radiology, and BRIC, IDEA Lab, University of North Carolina at Chapel Hill, North Carolina. dgshen@med.unc.edu.

Keywords

brain disease diagnosis; Alzheimer's disease; mild cognitive impairment (MCI); hierarchical classification; local patch; SVM classifier

INTRODUCTION

Brain images such as magnetic resonance images (MRI) are providing a powerful in vivo tool to help understand the disease-induced neural changes, due to Alzheimer's disease (AD) or its early stage such as mild cognitive impairment (MCI) [Davatzikos et al., 2010; Hinrichs et al., 2009; Leung et al., 2010; Li et al., 2011; Magnin et al., 2009; Mueller et al., 2005; Querbes et al., 2009; Tandon et al., 2006; Wolz et al., 2011; Zhang and Shen, 2011]. Recently, various classification methods have been proposed to identify the changes related to brain diseases and further decode the disease states by using neuroimaging data [Cuingnet et al., 2011; Davatzikos et al., 2008a; Fan et al., 2005; Hinrichs et al., 2009; Magnin et al., 2009; Oliveira et al., 2010; Wolz et al., 2011]. In most of these classification methods, two main steps are usually involved, i.e., (1) extraction and/or selection of discriminative features from the neuroimaging data, and (2) design of a supervised classifier for performing classification. Details for these two steps are briefed below.

The original brain images are usually too large and noisy to be directly input into the classifier for classification, and more importantly not all image information is useful for classification. Thus, feature extraction is necessary and important for extracting more relevant and discriminative features for neuroimage analysis and classification. In general, three types of MR imaging features were often extracted to detect the abnormal brain structures with AD, which include tissue densities (e.g., gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF)), cortical thickness, and both shape and volume of certain structure such as hippocampus [Cuingnet et al., 2011]. The volume in the region of interest (ROI), labeled by warping of a pre-labeled atlas, was often used to investigate brain abnormality [Lao et al., 2004; Magnin et al., 2009]. However, since the atlas-based ROI parcellation may not adapt well to the diseased-related pathology, the abnormal region may be part of one ROI or span over multiple ROIs, which may affect the feature discriminability. To address this issue, Fan et al. [Fan et al., 2007] proposed to adaptively partition the brain image into a number of most discriminative brain regions according to a predefined similarity measure, and then extracted regional features for brain disease classification. Although the ROI-based methods can significantly reduce the feature dimensionality and are robust to noise and registration error, the ROI-based regional features are generally very coarse and thus not sensitive to detect the small changes related to brain diseases. This limitation could be potentially alleviated by the voxel-wise analysis methods, i.e., using voxel-wise imaging features to identify the small brain abnormality [Baron et al., 2001; Ishii et al., 2005]. On the other hand, it was observed that the disease-induced structural changes also occur in several inter-related-regions, thus the correlations between different brain regions could also be extracted for more accurate characterization of brain pathology [Zhou et al., 2011]. By considering the nature of these pathological changes, the rich features related to both local brain regions and the inter-brain regions can be extracted

from brain images to avoid missing the important characterization of disease pathology. However, if more irrelevant and noisy information are included in the feature set, the disease classification and interpretation could become very difficult due to the small number of training samples in the neuroimaging study. For example, the support vector machines (SVM) classifier, which is often used for classification of brain disease [Fan et al., 2007; Klöppel et al., 2008; Magnin et al., 2009; Zhang et al., 2011], experiences a notable drop in classification accuracy when the number of irrelevant and noisy features is extremely large [Chapelle et al., 2002].

To address the above problem, principle component analysis (PCA) [Jolliffe, 2005] is popularly used to perform linear transformation of the data into a lower dimensional feature space for reduction of feature dimensionality [Davatzikos et al., 2008b; Yoon et al., 2007]. However, the main problem related to PCA is that the feature extraction is done independently from the subsequent classification task, thus potentially affecting the final classification results. Another popular solution to the above problem is to select the most discriminative features and eliminate the redundant features for further reduction of feature dimensionality and improvement of classification performance [Chu et al., 2012; Davatzikos et al., 2008a; Fan et al., 2005; Zhou et al., 2011]. For example, Chu et al. have compared four different feature selection methods followed by a linear SVM classifier [Chu et al., 2012]. Their experimental results show that feature selection does improve the classification accuracies, but it depends on the method used. Since the disease-induced brain structural changes often happen in the local focused regions, rather than isolated voxels, the local spatial contiguity of the selected features should be carefully considered for achieving better classification performance. For this purpose, the neighboring voxels with discriminative features (identified by feature selection) were jointly used for classification [Vemuri et al., 2008]. Although promising results have been reported for brain image analysis in the above studies, it is still potentially advantageous to investigate building and combining multiple classifiers for making full use of the rich imaging and structural information, for improved classification performance.

In this article, we propose a novel classification framework for analysis of voxel-wise neuroimaging data based on hierarchical fusion of neuroimaging features and decisions of multi-level classifiers in a layer-by-layer and local-to-global fashion. The spatially normalized grey matter (GM) of each T1-weighted MR brain image is computed as imaging features for classification. Note that the hierarchical classification framework was often used to solve the complex problem by gradually decomposing it into a number of easier-to-solve tasks [Scalzo and Piater, 2007; Singh et al., 2008]. A hierarchical generative model was also proposed to model the spatial relations and high-level appearance between correlated features and further generate the adaptive patch features to a SVM classifier for object classification in [Scalzo and Piater, 2007]. In addition, a hierarchical feature fusion model was proposed to combine feature fusion and decision fusion in [Scalzo et al., 2008]. Different from all these methods, we propose a hierarchical classification method that builds multilevel classifiers with supervised learning to gradually integrate imaging and spatial-correlation features for more accurate classification. The individual classifiers at the same level evaluate the classification abilities of the imaging features in different brain regions. On the other hand, the high-level classifiers work on larger brain regions than the low-level

classifiers. Figure 1 shows the hierarchical structure of our proposed classification framework. Specifically, the whole brain image is first partitioned into a number of local three-dimensional patches, with each containing only a subset of whole feature space. Second, for each patch, two different low-level classifiers are specially built with the use of local imaging and spatial-correlation features, respectively. Third, instead of directly combining low-level classifiers to make a final decision, the classifier outputs and the statistical imaging features at different brain regions are further integrated into a feature vector to construct the respective high-level classifiers for classification. Finally, the classification outputs of all high-level classifiers are combined to make the final classification.

The main contributions of this article can be summarized as follows. (1) A classification framework based on the hierarchical fusion of multi-level classifiers is proposed to gradually transform the high-dimensional imaging and structural data into more and more compact representations. Thus, this large-scale classification problem is hierarchically decomposed into a set of easy-to-solve small-scale problems, which is expected to improve the classification performance. (2) The imaging and spatial-correlation features of the whole brain image are extracted and gradually integrated into a hierarchical framework for more efficient and accurate classification. (3) The local spatial contiguity of image features is greatly respected in classification by using a hierarchical spatial structure that is built from small local patches to larger brain regions. It is worth noting that this article is the extension of our recently published workshop article [Liu et al., 2012], with some major differences as listed next. First, more detailed description and illustration of our method are provided in all sections, for allowing other people to better understand our method. Second, in the result section, we have also provided the new classification results between NC and MCI, the comparison with existing methods (Comparison With Existing Methods section), the list and analysis of the top-selected brain regions (Top Selected Regions section), and the discussion (Discussion section).

The rest of this article is organized as follows. Method section presents the details of the proposed classification framework. In Results section, experiments are presented to demonstrate the classification accuracy and the advantage of the proposed method. Finally, we conclude this article and discuss the possible future directions in Conclusion section.

METHOD

In this section, we will present our proposed classification algorithm. Figure 2 shows the flow chart of the proposed hierarchical classification algorithm by gradual fusion of multilevel classifier decisions and features. In the proposed method, the low-level classifiers are used to respectively transform the imaging and spatial-correlation features of a local patch, with supervised learning, into more compact representations (such as in certain intermediate feature spaces). Then, the outputs of these low-level classifiers together with the coarse-scale imaging features (i.e., statistical measures of local patches) are integrated to build the high-level classifiers, with each evaluating the features in different large brain regions. Finally, the classification is performed by ensemble of the outputs of all high-level classifiers. Accordingly, the proposed hierarchical classification framework can be divided

into five main steps: Imaging features, patch extraction, construction of low-level classifiers, construction of high-level classification, and final classification, as detailed one by one in the next.

Imaging Features

Although the proposed classification framework makes no assumption on a specific neuroimaging modality, for demonstrating its performance, the T1-weighted MR images, which have been widely used for detection of AD and MCI in the past decades, are used in this work. Before extraction of the imaging features for classification, pre-processing of these brain images is performed for reliable feature extraction. Specifically, all T1-weighted MR brain images are first skull-stripped and cerebellum-removed after a correction of intensity inhomogeneity [Sled et al., 1998]. Then, each brain image is segmented into three brain tissues, e.g., gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Finally, all the three tissues of each brain image are spatially normalized onto a standard space by a mass-preserving deformable registration algorithm [Shen and Davatzikos, 2003]. Since GM is more related to AD and MCI than the WM and CSF, the spatially-normalized GM volumes, which are also called as GM tissue densities, are used as the imaging features for classification in this work. The voxel-wise GM densities describe both GM information in the original subject and also its local geometric deformation relative to the selected brain template.

Patch Extraction

Given a GM tissue density map and patch size $w \times w \times w$, a simple way to extract the local patch is to uniformly divide the whole-brain image into a number of three-dimensional patches. However, this partition does not consider the different discriminability of individual brain regions and thus is not optimal for extraction of the discriminative local patches. To alleviate this problem, we first perform the t-test on each voxel with the training image set and select the voxels with significant group difference (i.e., with the resulting P value of t-test smaller than 0.05) among all the voxels in the image space. Second, for each selected voxel, we compute the mean of the P values in its local neighborhood of size $w \times w \times w$ and then use the mean P value to sort all selected voxels in ascending order. The first three-dimensional patch of size $w \times w \times w$ is extracted to be centered at the voxel with the smallest mean P value, followed by the second local patch centered at the voxel with the second smallest mean P value. Note that each new extracted patch should have less than 50% overlap with the previous extracted patches, i.e., the distance between the centers of any two patches is larger than $w/2$. Finally, by repeating the above step until all selected voxels are visited, we can obtain a set of three-dimensional local patches with size $w \times w \times w$, i.e., totally K patches denoted as $P = \{P_1, \ldots, P_k, \ldots, P_K\}$.

Construction of Low-Level Classifiers

For each extracted local patch P_k, we build two low-level classifiers $C_{1,k}$ and $C_{2,k}$ based on different types of low-level features related to the patch. Specifically, to capture both imaging and structural information from the neuroimaging data, two types of features are extracted for each patch, i.e., the local imaging features (GM densities of each patch) and the

correlations between pairs of local patches (namely spatial-correlation features in this article), respectively. Instead of building a single low-level classifier by fusion of these two types of features, we propose to build an independent classifier with each type of features. In this way, the difficult task in classification of high-dimensional features is now divided into a number of easier-to-do classification tasks using much lower-dimensional features. Then, based on the respective low-level imaging features, the classifier $C_{1,k}$ can be constructed with a supervised learning method such as SVM.

The second low-level classifier $C_{2,k}$ is constructed for each patch based on the spatial-correlation features, which characterize the relationship between different patches of the same subject and thus can capture more rich information about the pathology of AD and MCI. To do this, each patch is first represented by a feature vector that consists of the GM densities in that local patch. Then, the interaction between two patches within the same subject is computed as the Pearson correlation coefficient of their corresponding feature vectors (consisting of the GM densities in patches), to measure the similarity of imaging features between a pair of patches. When a patient is affected by AD or MCI, the correlation value of a particular brain patch with another patch will be potentially affected due to some factors such as atrophy. It is worth noting that the correlation provides a second-order measurement of the GM densities of the local patches. As a higher-order measurement, this new feature is more descriptive [Zhou et al., 2011] and can provide complementary information different from local GM densities for classification. Considering that the correlations can be computed between any pair of local patches in each subject, the feature dimensionality of all correlations is $K \times [K - 1]/2$, which is usually larger than 5,000. This will make it difficult to train a single classifier. Thus, for efficient training, we build a low-level classifier for each patch by using the correlations of this patch with all other patches in the same image, and thus obtain K low-level classifiers finally. Therefore, for all K extracted patches, we can obtain totally $2K$ low-level classifiers, which can be denoted as $C_i = \{C_{i,1}, \ldots, C_{i,k}, \ldots, C_{i,K}\}$, $i = 1, 2$.

High-Level Classification

After constructing two sets of low-level classifiers $C_i = \{C_{i,1}, \ldots, C_{i,k}, \ldots, C_{iK}\}$, $= i = 1, 2$. from all K local patches, a common ensemble approach is to directly combine the classifier outputs to make the final classification. However, since the low-level classifiers are built with the features from local patch, their performances may be limited especially when the affected brain regions are larger than the local patch. The limited performance of the low-level classifier will affect the final ensemble classification performance. To alleviate this problem, we propose to combine the decision outputs of low-level classifiers, along with the coarse-scale imaging features in each local patch, to build multiple high-level classifiers placed at different larger brain regions. The high-level classification results on different brain regions are finally combined to make the final classification.

High-level features—We combine three types of high-level features to build each high-level classifier at a specified brain region as described below. The first two types of high-level features are the outputs of two low-level classifiers $C_{1,k}$ and $C_{2,k}$ in each local patch k. Instead of using the class label, the output of each low-level classifier is computed as the

continuous value that evaluates the probability of each patch belonging to different classes. For example, it can be treated as an estimate of the class posterior probability as used in Tu and Bai [2010]. Specifically, for the SVM classifier that outputs the relative distance to the decision boundary (i.e., the signed distance), we can use a logistic function to convert the classifier output into a probability, belonging to [0 1]. The classifier output can be considered as the patch-level representation of the low-level imaging features and is thus more relevant to class label. With the supervised learning, the high-dimensional imaging feature space is now transformed into the compact high-level feature space. Therefore, we can treat the outputs of low-level classifiers as the inputs for high-level classification in the certain intermediate feature space.

The third type of high-level features is the statistical measures that are computed as the mean and standard deviation of the GM densities in each local patch. Although these statistical features capture the coarse-scale imaging information with limited discriminative information, they can achieve higher robustness to noise and thus are useful for high-level classification, as will be demonstrated in the experimental results. All these three types of features computed from the local patches in a specified brain region will be concatenated into a feature vector to train the high-level classifier that is responsible for the specific large brain region, as detailed below. In addition, for each feature f_i in the training samples, it will be normalized using equation $f_i = \frac{f_i - \bar{f_i}}{\sigma_i}$, where $\bar{f_i}$, and σ_i are the mean and standard deviation of the feature f_i across all training samples. Similarly, this normalization process will be applied to the corresponding feature of each test sample.

High-level classifiers—The information related to the disease can be distributed over some distant brain regions with arbitrary shape and size. To maximize the prediction accuracy, the high-level classifier should be instantiated at the informative brain regions, i.e., through coarse subdivision of the brain volume. Similar to the process of extracting three-dimensional local patches as described above, local patches can be agglomerated to form the highly-informative brain regions with respect to the disease classification, and then the high-level classifiers can be constructed to maximize the discriminability of the high-level features in each brain region. Also as introduced below, the size of each brain region can be optimized to achieve the best performance for its respective high-level classifier, and the obtained sizes can be different across different brain regions.

Specifically, we first perform the cross-validation test on the low-level classifiers with the training data, and thus obtain their classification accuracies, which evaluate the classification abilities of the local patches. Then, the local patches are sorted according to the classification accuracies of the low-level classifiers constructed with the local GM densities, and the first brain region is centered at the local patch with the highest classification accuracy. To obtain the most informative brain region, we first change the size of brain region (i.e., the radius from the first selected patch) within a predefined range, and then use all high-level features within this brain region to train a high-level classifier. In the meanwhile, the classification performance will be cross-validated for each size of brain region. Finally, we can select a brain region that can yield the highest classification accuracy, and place the first high-level classifier on this brain region. After building the first

high-level classifier, the second and other subsequent high-level classifiers are similarly constructed one by one, with their respective brain regions partially overlapped by less than 50%. Finally, we can obtain a set of M high-level classifiers trained with the high-level features in M different brain regions, i.e., HC = {HC_1, ..., HC_j, ..., HC_M}. It is worth noting that each brain region for building the high-level classifier may not have the same number of patches or the same number of features.

It is worth noting that our proposed hierarchical classification framework is not limited to any particular choice of classifier model. Many state-of-the-art classifiers, such as SVM and linear discriminant analysis (LDA), can be used to build the base classifiers. In this work, we choose the linear SVM classifier without any threshold as the base classifier to build both the low-level and high-level classifiers. For simplicity, we implement the SVM classifier with a linear kernel by using the SVM functions provided by MATLAB software [Kecman, 2001]. The value of the box constraint C for the soft margin is set to the default value 1, and we also use the SMO Method to find the separating hyperplane.

Final (Ensemble) Classification

The final classification is made by combination of the *M* high-level classifiers with the weighted voting strategy. In general, the ensemble classifier by fusing multiple classifiers is superior to the single classifier when the predictions of component classifiers have enough diversity [Brown et al., 2005]. In our case, the multiple high-level classifiers are trained with the features of different brain regions, thus giving a certain degree of diversity to improve the ensemble classification. However, since we allow the overlap among different brain regions selected for high-level classifiers, the discriminating capabilities of some high-level classifiers may be similar to some extent. In addition, the disease-related pathological changes often happen in a small number of brain regions. Thus, it is important to select a subset of high-level classifiers with larger discriminating capability for more accurate ensemble classification and also for facilitating the interpretation of classification results. Although the exhaustive search of all possible classifier combinations allows obtaining the optimal subset of high-level classifiers for final ensemble, it is computationally expensive when the number of high-level classifiers is large. Greedy approach focuses on adding or removing a specific classifier at each time for maximizing the improvement in the ensemble performance (Ruta and Gabrys, 2005), thus taking less computational cost with good performance. In this article, we employ a forward greedy search strategy to select an optimal subset of high-level classifiers for final fusion, as described in Figure 3.

It is worth noting that the classifier selection in the above is performed on the training set and thus may not be optimal for the testing set. To improve the generalization, we divide the training set into 10 folds, and in each fold a subset of classifiers is selected using the forward greedy search. The selection frequency of each classifier is computed over all folds, with the high frequency indicating the high likelihood of the respective classifier to improve the ensemble accuracy. Thus, the selection frequency of each high-level classifier is treated as its weight in the final voting. Specifically, for a test sample *x*, the weighted sum of the prediction outputs of M high-level classifiers is used to make the final classification:

$$D(x) = \text{sign}\left(\sum_{j=1}^{M} w_j PC_j(x)\right) \quad (2)$$

where $PC_j(x)$ is the prediction output of the j-th high-level classifier $PC_j(x)$ for the test sample x, and is the respective weight assigned to the j-th high-level classifier which is computed as the selection frequency of the j-th high-level classifier in the 10-fold testing of the forward greedy search method.

RESULTS

Data Set

The data used for evaluation of our proposed hierarchical classification algorithm were taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (available at: www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations, as a $60 million, 5-year public–private partnership. The primary goal of the ADNI has been to test whether serial magnetic resonance imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California, San Francisco. ADNI was the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. The study subjects were recruited from over 50 sites across the US and Canada. They gave the written informed consent at the time of enrollment for imaging and genetic sample collection and completed the questionnaires approved by the Institutional Review Board (IRB) of each participating site.

Our experimental evaluations are based on a portion of the ADNI database. We use the T1-weighted MR imaging data from the baseline visit. MRI acquisitions have been done according to the ADNI acquisition protocol in [Jack et al., 2008]. T1-weighted MR image data from 652 ADNI participants are used for evaluation in the experiments. These 652 subjects include 198 AD, 225 MCI (including 112 stable MCI (sMCI) and 113 progressive MCI (pMCI)), and 229 NC. Table I presents a summary of the demographic characteristics of the studied subjects (including the number, age, gender, and MMSE of the subjects).

The image processing of the T1-weighted MR brain images was performed as described in Imaging Features section. The spatially normalized GM tissues, i.e., GM densities (with isotropic voxel size of $1 \times 1 \times 1$ mm^3), are used as the imaging features. To reduce the impact of noise, registration error, and inter-individual anatomical variations, the tissue density maps were further smoothed using a Gaussian filter (with a sigma value of 1.0) and

then down-sampled by a factor of 4 for saving the computational time and memory cost, without sacrificing the classification accuracy as confirmed by our experiment. Final imaging size was of $64 \times 64 \times 64$ voxels, with the voxel size of $4 \times 4 \times 4$ mm³. To build the low-level classifiers, we partitioned the GM density map into a number of three-dimensional local patches. For simplicity, the patch size was set to $11 \times 11 \times 11$ in the whole image, although it could be adaptively determined in different brain regions. In total, we obtained more than one hundred of patches for each brain image. Then, two low-level classifiers were built for each patch by using the GM densities and spatial-correlation features, respectively. In our experiments, to evaluate the classification performance, we use a 10-fold cross-validation strategy to compute the classification accuracy (ACC), which evaluates the proportion of correctly classified subjects among the whole test population. In addition, we also compute the sensitivity (SEN), i.e., the proportion of AD (or MCI) patients correctly classified, and the specificity (SPE), i.e., the proportion of correctly classified normal controls for further evaluation. In each time, one fold of the data set was used for testing, while the other remaining nine folds were used for training. The training set was further divided into 10 folds to optimize the parameters in our method, which include the size of the brain region in building the high-level classifiers and the weight assigned to each high-level classifier in final ensemble.

Classification Results

We conducted two experiments to investigate the effectiveness of the proposed hierarchical classification algorithm in identification of AD (or MCI) from normal controls. The first experiment was performed to test the efficacy of different features used for classification. In general, three types of features, i.e., GM-density-based classifier outputs (GCO), correlation-based classifier outputs (CCO), and statistical measures (SM), are extracted from each patch to build the high-level classifiers. To evaluate the efficacy of these features, we computed the classification accuracy by ensemble of the high-level classifiers built with different features. When using GCO or CCO features, or their combinations, usually 20 to 30 high-level classifiers were obtained for final ensemble. Since the statistical measure (SM) is a kind of coarse-scale feature with limited discrimination in small brain region, we obtained only one SM-based high-level classifier, instantiated on a relatively large brain region. The classification results with respect to the use of different features and their combinations are summarized in Tables II and III for classification of AD versus NC and MCI versus NC, respectively.

As we can see from these results, the statistical measures (SM) have limited information and thus result in low classification accuracy when only this type of features is used for classification. However, the SM can improve the classification accuracy when combined with other two types of features. In addition, GCO are generated based on the GM densities of each local patch, while CCO are generated based on the correlations between pairs of patches. These two types of features can provide complementary information for classification, and thus the combination of GCO and CCO improves the classification accuracy. Finally, combination of these three types of features by our proposed hierarchical method can further improve the classification performance.

The second experiment was conducted to test the effectiveness of the hierarchical fusion used in the proposed classification framework. Specifically, we compared the performance of the proposed hierarchical fusion method with those of other two possible classification methods. The first possible classification method is to build a single global classifier for final classification, by using PCA to reduce the dimensionality of GM density features. In particular, the dimensionality of the reduced feature space by PCA was optimally determined via the 10-fold cross-validation using the training set. We changed the feature dimensionality within a predefined range (i.e., from 10 to the number of training samples) and selected the dimensionality with the minimum classification error rate for the test set. The PCA-based classification method is used as an example of conventional classification method with a single global classifier, to be compared with our proposed method that uses the ensemble of multilevel classifiers. The second classification method is to directly ensemble the decisions of both the GM-density-based and correlation-based low-level classifiers using the weighted voting. The weight assigned to each individual classifier is determined using the same strategy as used in the proposed hierarchical classification method (Final (Ensemble) Classification section). The classification results and their comparisons with respect to different methods are summarized in Tables IV and V for classification of AD versus NC and MCI versus NC, respectively. Also, their receiver operating characteristic (ROC) curves for classification of AD versus NC and MCI versus NC are given in Figures 4 and 5, respectively. These results demonstrate that both classification methods by fusion of multiple classifiers can achieve better performance than the single classifier. Moreover, the proposed hierarchical fusion method can further improve the classification performance compared with the direct fusion method. Specifically, the proposed hierarchical classification can achieve a high accuracy of 92.0% (with sensitivity of 91.0% and specificity of 93.0%) for classification of AD versus NC and 85.3% for classification of MCI versus NC, respectively.

For interpreting our results, we further visualize the patches that are selected for disease classification using GCO features. Since a linear SVM classifier is used in our experiment, the classifier weights are used to show the contribution of each voxel in the selected patches. Figure 6a,b show the weight maps for the voxels in the selected patches when used for AD versus NC and MCI versus NC classifications, respectively.

On the other hand, to illustrate the contributions of the correlation classifier (CCO), we show in Figure 7a three-dimensional graph overlaid on a three-dimensional transparent brain, which is generated with the BrainNet Viewer package (available at: http://www.nitrc.org/projects/bnv/). The nodes of the graph denote the selected patches, while the thickness of the graph edge indicates the weight of the linear SVM classifier built with the respective patch correlations. For better illustration, the important patches and classifier weights are selected as follows. First, we sort the correlation classifiers in ascending order of their classification accuracies and select top 10 classifiers. Then, for each selected classifier, the patches with 10 highest classifier weights are selected as the important patches. Finally, all selected patches are used to generate the graph nodes, and the corresponding classifier weights are used to generate the graph edges. Figure 7a,b show the three-dimensional graphs for the cases of AD versus NC and MCI versus NC classification, respectively.

Comparison with Existing Methods

Furthermore, we compare the results of the proposed classification method with some results recently reported in the literature, which were also obtained based on the baseline MRI data of ADNI subjects. In particular, four recent classification methods for AD and MCI diagnosis are compared as briefly described in the following.

- In [Hinrichs et al., 2009], the linear program (LP) boosting method with novel additional regularization was proposed to incorporate the spatial smoothness of MR imaging space into the learning process and improve the classification accuracy. Only classification results for AD versus NC were provided in this article.

- In [Cuingnet et al., 2011], 10 methods on different types of MRI-based features, which included five voxel-wise imaging feature based methods, three cortical thickness based methods, and two hippocampus based methods, were compared for classification of AD versus NC and MCI versus NC with the linear SVM classifier. The best results, which were obtained using voxel-wise GM densities, were provided for comparison in our article.

- In [Zhang et al., 2011], 93 volumetric features were extracted from the 93 regions of interest (ROI) in GM densities for both AD and MCI classification and a single SVM classifier was constructed to make classification.

- More recently, four types of MRI-based features, i.e., hippocampal volume, tensor-based morphometry, cortical thickness, and manifold-learning based features, were combined to achieve improved classification accuracies with both linear discriminant analysis (LDA) and SVM classification approaches in [Wolz et al., 2011]. For comparison, we present their best results that were obtained with the LDA classification approach.

The classification results of the above four methods, along with our proposed method, for classification of AD versus NC and MCI versus NC are summarized in Tables VI and VII, respectively. These results further validate the efficacy of our proposed classification method.

It is worth noting that, in Tables VI and VII, we just list the results of different methods reported in the literature. These different methods may not use exactly the same subjects from ADNI, and thus the comparison of their results needs to be careful. In the following, we compare the performance of our proposed with a specific method reported in the literature, by using the same dataset. In [Chu et al., 2012], the impact of sample size and feature selection on brain classification were extensively studied by using the GM features and SVM classifier. In particular, they compared four different feature selection methods, i.e., one prior-knowledge based method, two data-driven methods, and one hybrid method. Their experimental results showed that the most accurate classification was achieved by feature selection using the prior knowledge about the regions of brain atrophy found in previous studies, i.e., using all GM voxels in the hippocampal and parahippocampal mask. Therefore, this prior-knowledge based method is used here for comparison. Specifically, by using our template with 93 manually-delineated ROIs [Kabani et al., 1998], we can label all GM voxels in the hippocampal and parahippocampal regions for each subject, and then we

can use a linear SVM for classification. For fair comparison, the same training and testing data sets are used as our method. This prior-knowledge based method produces classification accuracy (ACC) of 84.5%, along with SEN of 82.3% and SPE of 86.4%, for classification of AD versus NC. A simple paired t-test of the accuracies in each fold of cross validation was also performed to test the difference between our method and this prior-knowledge based method. The obtained *P* value, 1.7954e-004, indicates that our method is statistically better than the prior-knowledge based method.

Top Selected Regions

For better understanding of the brain regions selected by the proposed method for AD or MCI classification, we picked out the most discriminative patches according to the hierarchy of classifiers based on the training data set. Since our proposed hierarchical classification method builds multilevel classifiers on different brain regions, the classification accuracy of the respective classifier indicates the importance of the corresponding brain region in classification. Specifically, we first selected the brain region related to the high-level classifier with the highest accuracy. From the selected brain region, we then selected the local patches that give high accuracy with respect to the GCO and CCO in the low-level classification. It is worth noting that the patch selection is performed on the training data only. Thus, the selected patches at each cross-validation fold may be different. For example, we checked the selected patches from all cross-validation folds, and found that some selected patches do vary across different folds. Thus, we compute the frequencies of the voxels included in the selected patches in all folds. For illustration purpose, we generated the frequency maps for the voxels in the selected patches during the cross-validations for AD and MCI classifications (see Fig. 8). It can be observed from Figure 8 that the most-affected regions detected by our classification method include hippocampus, parahippocampal gyrus, entorhinal cortex, and amygdala, which are consistent with those reported in the literature for AD and MCI studies [Cuingnet et al., 2011; Hinrichs et al., 2009; Zhang et al., 2011].

Discussion

In this article, we have proposed a hierarchical classification framework to gradually combine features and classifier decisions into a unified multilevel model for analysis of MR images, to assist the diagnosis of AD and MCI. Different from the conventional classification methods that build a single classifier with all input features, the proposed method divide the difficult task for classification of high-dimensional features into many low-dimensional classification problems that are easier to solve. The rich imaging and spatial-correlation features of the whole brain image are extracted and gradually integrated into a hierarchical framework for more efficient and accurate classification. Our experimental results show that the hierarchical fusion of these two features can improve the classification performance. To the best of our knowledge, there are no previous studies on combining these two features for classification. More importantly, the local spatial contiguity of imaging features is greatly respected in classification by using a hierarchical spatial structure that is built from small local patches to larger brain regions. This strategy can make better use of the local information than the ROI-based methods [Fan et al., 2007;

Zhang et al., 2011]. Ensemble learning is a kind of machine learning technique by combining multiple weak classifiers to build a strong classifier. Adaboost learning is a popular ensemble learning by aggregating base classifiers to successively estimate their errors and focusing more and more on the instances misclassified by previous classifiers. However, AdaBoost is sensitive to noisy data and outliers since the noisy data will be also put with high weight in the subsequent classification, thus degrading the classification performance. The proposed hierarchical ensemble method is to aggregate multiple local classifiers gradually by fusing the multilevel classifier decisions and features into a global classifier, which is more robust to noisy data and outliers.

We also tested the proposed hierarchical classification method on the entire brain voxels, without limiting to the voxels with significant univariate group difference by *t*-test. The accuracies of the proposed method on the entire brain voxels are 91.35% (along with 89.92% SEN, 92.55% SPE) and 84.86% (along with 81.84% SEN, 87.79% SPE) for classifications of AD versus NC and MCI versus NC, respectively. These results further show that the proposed method is also robust to the size of feature space.

Selecting suitable patch size is important for achieving good classification performance. If the patch size is too small, each patch will have no enough information to offer good performance in the low-level classification, and also the number of patches or low-level classifiers will be too large which will significantly increase the computation cost in the classification. On the other hand, if the patch size is too large, more redundant or even confounding information will be included into each patch, which will affect the localization of informative brain regions and finally the ensemble classification result. To balance these, the patch size needs to be optimized, i.e., $11 \times 11 \times 11$ as we obtained, which leads to about 120 patches on average in our study. On the other hand, if other sizes are used, the classification performance may be affected.

As for the number of levels in the hierarchy to build the base classifiers, our current classification method adopts three levels of hierarchy. We have compared the experimental results with one, two, and three levels of hierarchy for classification in the second experiment. The experimental results show that the method with three levels of hierarchy performs better than others, because it can make better use of the local features and classifier decisions. However, when increasing the hierarchical levels to be more than three, we found that the classification performance is not further improved, but the computation complexity is increased.

CONCLUSION

In summary, we have presented a hierarchical classification algorithm for MRI-based diagnosis of AD and MCI in this article. To deal with the challenge of high-dimensional imaging features in the whole brain MRI during AD/MCI diagnosis, we proposed to gradually aggregate the low-level imaging features into the compact high-level representations via constructing multilevel classifiers with supervised learning. Thus, the large-scale classification problem with high-dimensional imaging features can be decomposed into a hierarchical set of small-scale classification problems, which are easier to

handle. In addition to the local imaging features, the spatial-correlations are also integrated into the hierarchical model for better classification. Experimental results on the baseline data of ADNI subjects show that the ensemble of multiple classifiers performs better than the single global classifier and importantly the hierarchical fusion of multi-level classifiers can further improve the classification performance.

Since different imaging modalities can provide complementary information for disease diagnosis, in the future work, we will extend our method to include other imaging features extracted from other modality of data. We will also investigate more advanced classifier ensemble method, i.e., sparse multiple kernel learning [Subrahmanya and Shin, 2010], for further improvement of the classification accuracy.

Acknowledgments

REFERENCES

Baron JC, Chetelat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. Neuroimage. 2001; 14:298–309. [PubMed: 11467904]

Brown G, Wyatt J, Harris R, Yao X. Diversity creation methods: A survey and categorisation. Information Fusion. 2005; 6:5–20.

Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Mach Learn. 2002; 46:131–159.

Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C. for the Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage. 2012; 60:59–70. [PubMed: 22166797]

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, Chupin M, Benali H, Colliot O. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage. 2011; 56:766–781. [PubMed: 20542124]

Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging. 2010; 32:2322. e19-2322.e27. [PubMed: 20594615]

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging. 2008a; 29:514–523. [PubMed: 17174012]

Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage. 2008b; 41:1220–1227. [PubMed: 18474436]

Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput Assist Interv. 2005; 3749:1–8. [PubMed: 16685822]

Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: Classification of morphological patterns using adaptive regional elements. IEEE Trans Med Imaging. 2007; 26:93–105. [PubMed: 17243588]

Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage. 2009; 48:138–149. [PubMed: 19481161]

Ishii K, Kawachi T, Sasaki H, Kono AK, Fukuda T, Kojima Y, Mori E. Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. Am J Neuroradiol. 2005; 26:333–340. [PubMed: 15709131]

Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJJLW, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008; 27:685–691. [PubMed: 18302232]

Jolliffe, I. Principal Component Analysis, Encyclopedia of Statistics in Behavioral Science. Vol. 3. New York: John Wiley & Sons, Ltd; 2005. p. 1580-1584.

Kabani N, MacDonald D, Holmes CJ, Evans A. A 3D atlas of the human brain. Neuroimage. 1998; 7:S717.

Kecman, V. Learning and Soft Computing-Support Vector Machines, Neural Networks, Fuzzy Logic Systems. Cambridge, MA: The MIT Press; 2001.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR, Ashburner J, Frackowiak RSJ. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008; 131:681–689. [PubMed: 18202106]

Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage. 2004; 21:46–57. [PubMed: 14741641]

Leung K, Shen KK, Barnes J, Ridgway G, Clarkson M, Fripp J, Salvado O, Meriaudeau F, Fox N, Bourgeat P. Increasing power to predict mild cognitive impairment conversion to Alzheimer's disease using hippocampal atrophy rate and statistical shape models. Med Image Comput Comput Assist Interv. 2010; 13:125–132. [PubMed: 20879307]

Li Y, Wang Y, Wu G, Shi F, Zhou L, Lin W, Shen D. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. Neurobiol Aging. 2011; 33:427. e15–27.e30. [PubMed: 21272960]

Liu, M.; Zhang, D.; Yap, P-T.; Shen, D. Hierarchical ensemble of multi-level classifiers for diagnosis of Alzheimer's disease. France. Proceedings of the Third International Workshop on Machine Learning in Medical Imaging (MLMI), Nice; 2012 Oct 1.

Magnin B, Mesrob L, Kinkingnehun S, Pelegrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehericy S, Benali H. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology. 2009; 51:73–83. [PubMed: 18846369]

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 2005; 1:55–66. [PubMed: 17476317]

Oliveira PJ, Nitrini R, Busatto G, Buchpiguel C, Sato J, Amaro EJ. Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease. J Alzheimer Dis. 2010; 19:1263–1272.

Querbes O, Aubry F, Pariente J, Lotterie JA, Demonet JF, Duret V, Puel M, Berry I, Fort JC, Celsis P. Early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve. Brain. 2009; 132:2036–2047. [PubMed: 19439419]

Ruta D, Gabrys B. Classifier selection for majority voting. Information Fusion. 2005; 6:63–81.

Scalzo, F.; Bebis, G.; Nicolescu, M.; Loss, L.; Tavakkoli, A. Feature fusion hierarchies for gender classification; Florida, USA. The 19th International Conference on Pattern Recognition (ICPR), IEEE, Tampa; 2008. p. 1-4.

Scalzo, F.; Piater, JH. Adaptive patch features for object class recognition with learned hierarchical models; Minnesota, USA. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Minneapolis; 2007. p. 1-8.

Shen D, Davatzikos C. Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. Neuroimage. 2003; 18:28–41. [PubMed: 12507441]

Singh R, Vatsa M, Noore A. Hierarchical fusion of multi-spectral face images for improved recognition performance. Information Fusion. 2008; 9:200–210.

Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998; 17:87–97. [PubMed: 9617910]

Subrahmanya N, Shin YC. Sparse multiple kernel learning for signal processing applications. IEEE Trans Pattern Anal Mach Intell. 2010; 32:788–798. [PubMed: 20299705]

Tandon R, Adak S, Kaye J. Neural networks for longitudinal studies in Alzheimer's disease. Artif Intell Med. 2006; 36:245–255. [PubMed: 16427257]

Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE Trans Pattern Anal Mach Intell. 2010; 32:1744–1757. [PubMed: 20724753]

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knop-man DS, Boeve BF, Petersen RC Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage. 2008; 39:1186–1197. [PubMed: 18054253]

Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, Soininen H, Lötjönen J. Multi-method analysis of MRI Images in early diagnostics of Alzheimer's disease. Plos One. 2011; 6:e25446. [PubMed: 22022397]

Yoon U, Lee JM, Im K, Shin YW, Cho BH, Kim IY, Kwon JS, Kim SI. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. Neuroimage. 2007; 34:1405–1415. [PubMed: 17188902]

Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage. 2011; 59:895–907. [PubMed: 21992749]

Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011; 55:856–867. [PubMed: 21236349]

Zhou L, Wang Y, Li Y, Yap PT, Shen D. Hierarchical anatomical brain networks for MCI prediction: revisiting volumetric measures. Plos One. 2011; 6:e21935. [PubMed: 21818280]
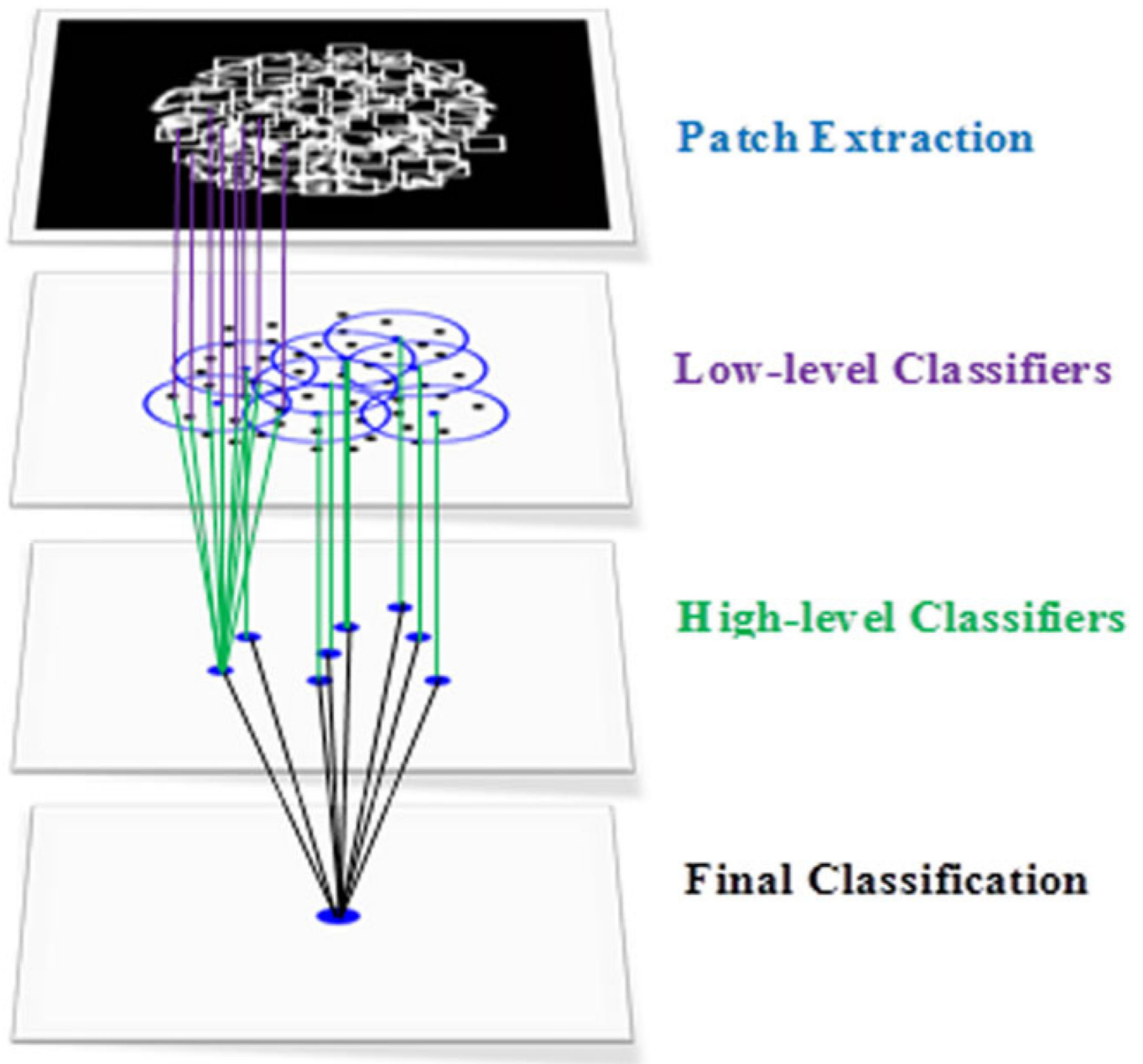
Figure 1.
The hierarchical structure of the proposed classification framework by using a two-dimensional slice for illustration, where the white squares denote local patches. The small, middle and large dots denote the low-level, high-level, and final classifiers, respectively, and the blue circles denote the brain regions where the high-level classifiers are placed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
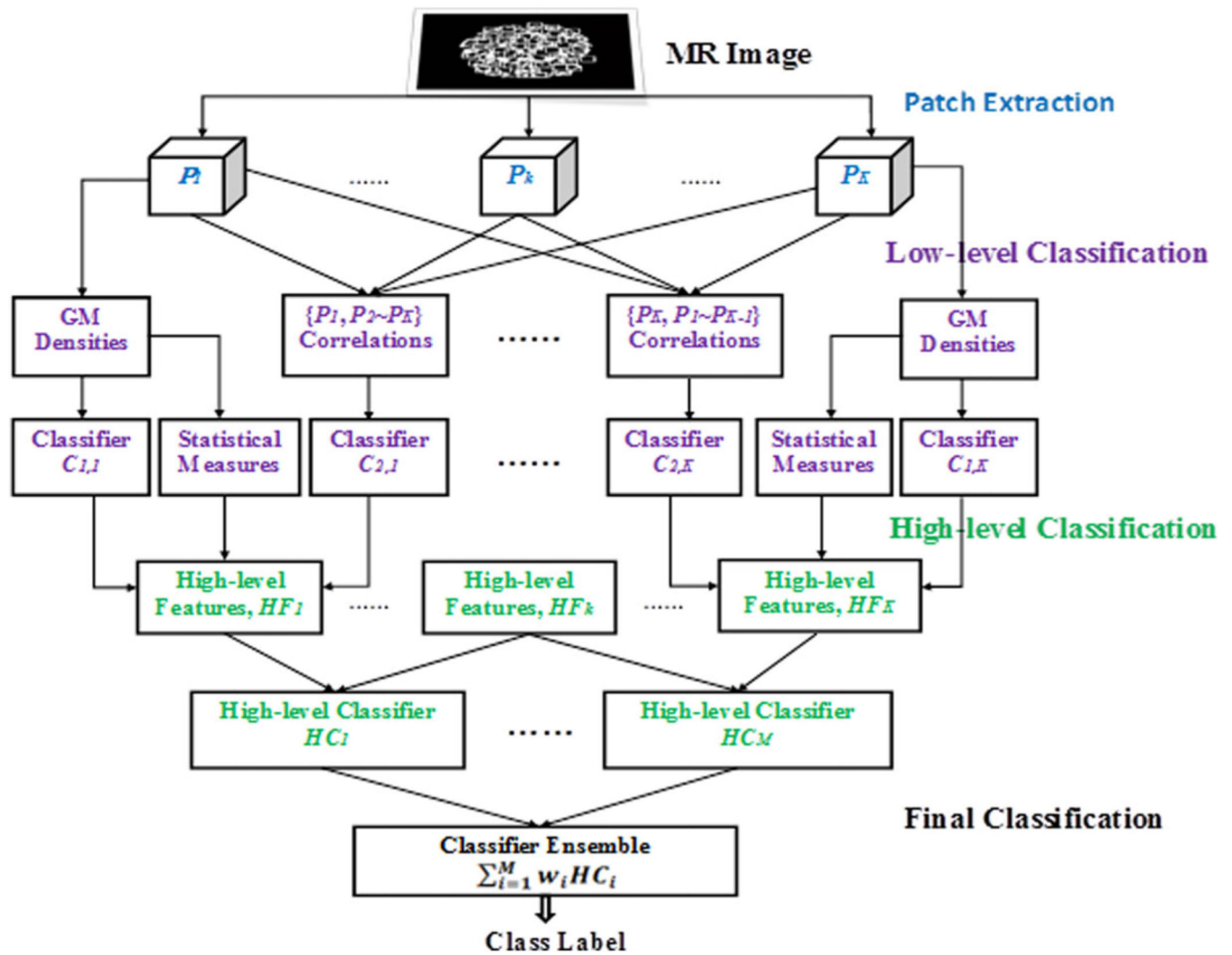
Figure 2.
The flow chart of the proposed hierarchical classification algorithm. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Classifier Selection with Forward Greedy Search

Input: Candidate Classifier Set: $CCS = \{HC_1, ..., HC_j, ..., HC_M\}$;

Selected Classifier Set: $SCS = \{\}$.

Output: Selected Classifier Set: SCS

Repeat: $t=1\sim M$

Step 1: For all $HC_j \in CCS$, evaluate the classification accuracy on the ensemble of classifiers $HC_j \cup SCS$, i.e., $Acc(HC_j \cup SCS)$.

Step 2: Select the classifier that maximally increases the accuracy of ensemble classification: $HC_k = \arg\max_{HC_j \in CCS} Acc(HC_j \cup SCS)$.

Step 3: Update SCS and CCS: $SCS = HC_k \cup SCS$, $CCS = CCS / \{HC_k\}$.

Step 4: If the combined classification accuracy cannot be further improved by adding any candidate classifier, stop the iteration. Otherwise, $t=t+1$ and go to Step 1.

Figure 3.
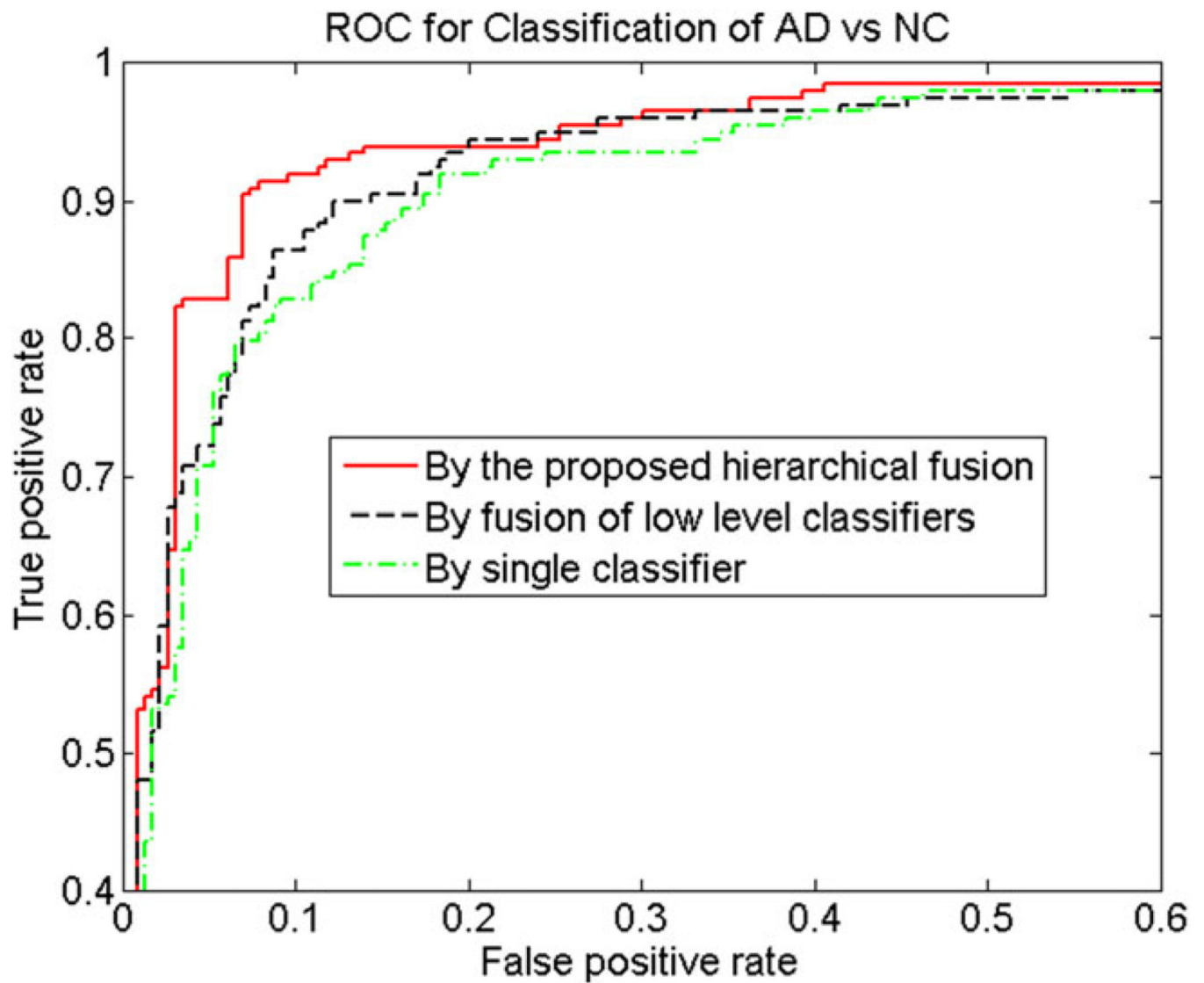Classifier selection with forward greedy search.

Figure 4.
ROC curves of three methods in AD vs. NC classification. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
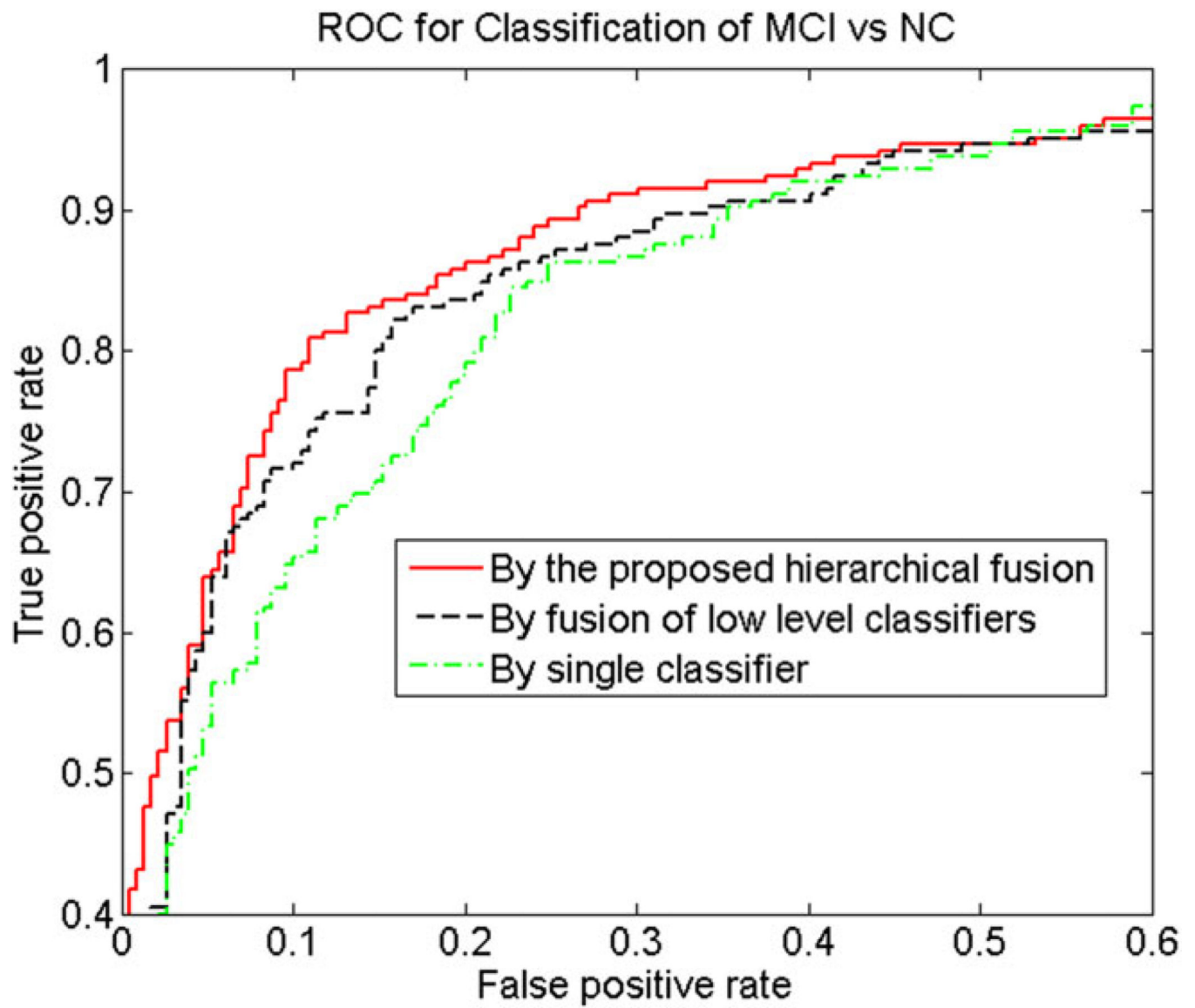
Figure 5.
ROC curves of three methods in MCI vs. NC classification. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
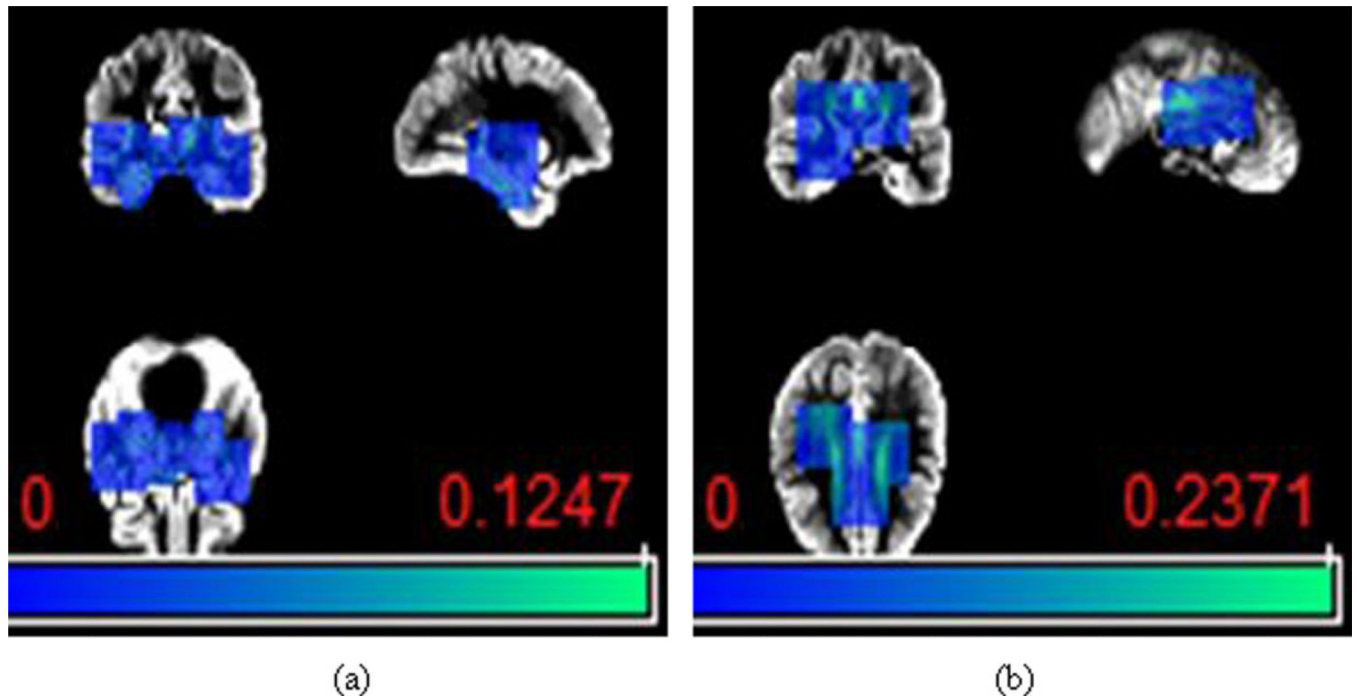
Figure 6.
The weight maps for the voxels in the selected patches for (**a**) AD vs. NC classification and
(**b**) MCI vs. NC classification. [Color figure can be viewed in the online issue, which is
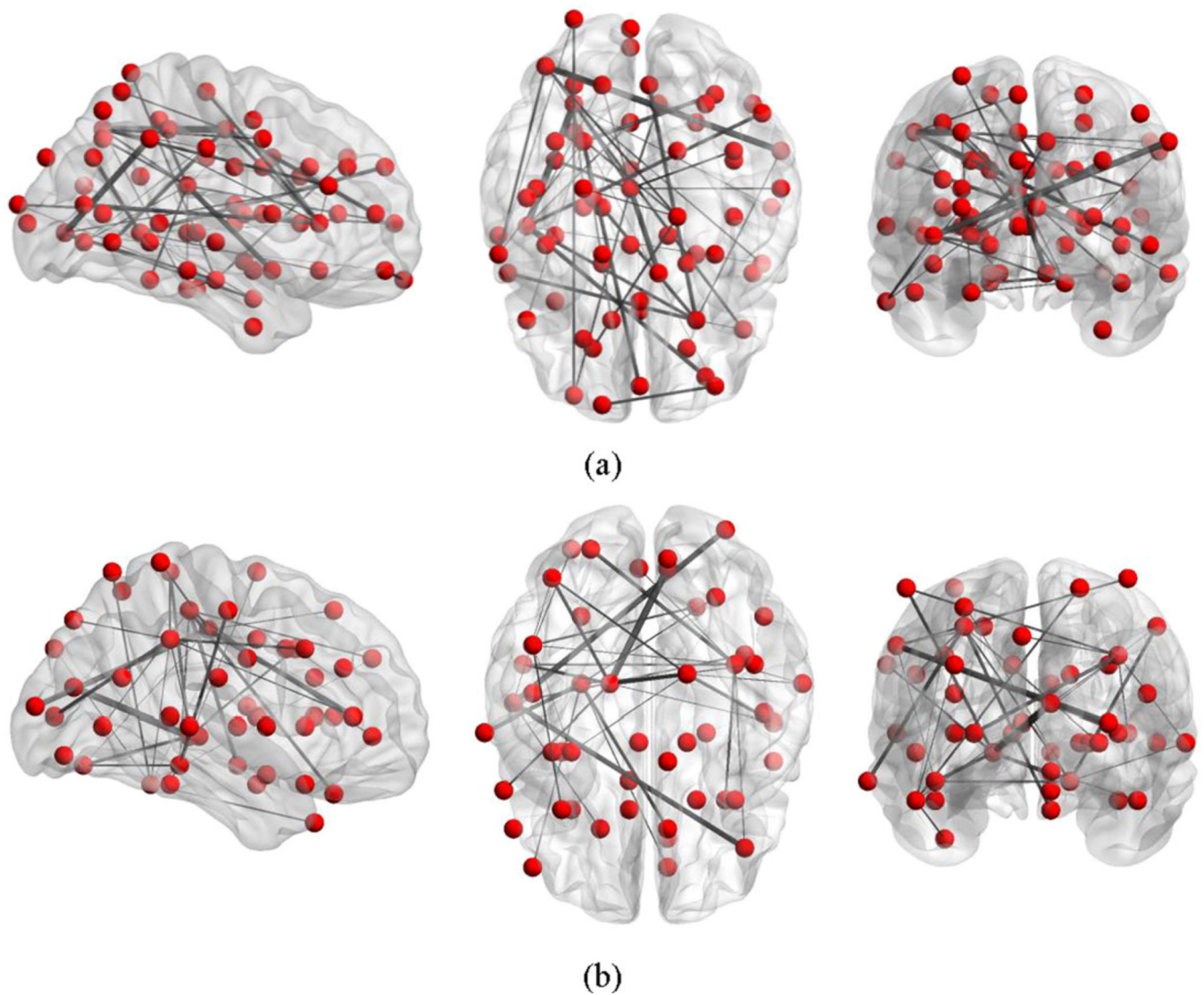available at wileyonlinelibrary.com.]

(a)

(b)

Figure 7.
The three-dimensional graphs generated with the nodes indicating the important patches and the edge thickness indicating the weight of the correlation classifier for (**a**) AD vs. NC classification and (**b**) MCI vs. NC classification. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
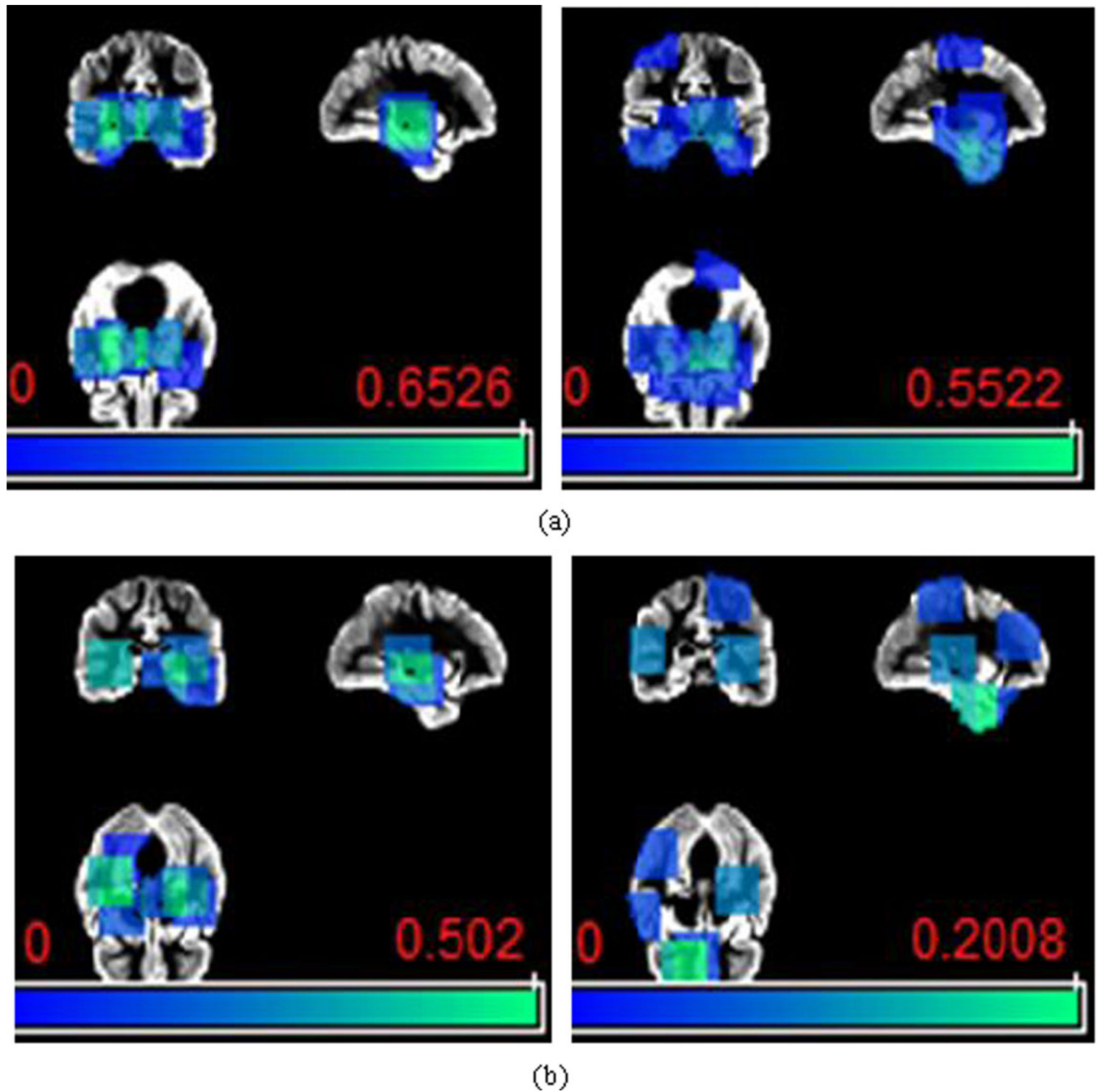
Figure 8.
The frequency maps for the voxels in the selected patches during the cross-validation for (**a**) AD classification and (**b**) MCI classification, with respect to the use of GCO (left) and CCO (right) features. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE I

Demographic characteristics of the studied subjects from ADNI database (denoted as mean ± standard deviation)

Diagnosis	Number	Age	Gender (M/F)	MMSE
AD	198	75.7 ± 77	103/95	23.3 ± 2.0
MCI	225	75.2 ± 7.4	154/71	26.7 ± 1.8
NC	229	76.0 ± 5.0	119/110	29.1 ± 1.0

TABLE II

Performance comparison for classification of AD versus NC on different features

Classification features	ACC (%)	SEN (%)	SPE (%)
Statistical measures (SM)	85.3	83.4	86.9
GM-density-based classifier outputs (GCO)	90.2	88.9	91.3
Correlation-based classifier outputs (CCO)	89.7	89.4	89.9
SM + GCO	91.1	89.5	92.5
SM + CCO	90.8	88.4	93.0
GCO + CCO	90.9	89.4	92.1
SM + GCO + CCO	92.0	90.9	93.0

TABLE III

Performance comparison for classification of MCI versus NC on different features

Classification features	ACC (%)	SEN (%)	SPE (%)
Statistical measures (SM)	74.1	73.4	74.7
GM-density-based classifier outputs (GCO)	83.7	80.1	87.3
Correlation-based classifier outputs (CCO)	82.5	81.8	83.0
SM + GCO	84.4	80.5	88.2
SM + CCO	83.2	81.0	85.2
GCO + CCO	84.2	82.3	86.0
SM + GCO + CCO	85.3	82.3	88.2

TABLE IV

Comparison of three methods for AD versus NC classification

Classification methods	ACC (%)	SEN (%)	SPE (%)	Area under ROC (%)
Single classifier	86.4	83.9	88.6	92.9
Direct fusion of low-level classifiers	89.7	86.9	92.1	93.9
Proposed hierarchical fusion	**92.0**	**91.0**	**93.0**	**95.2**

TABLE V

Comparison of three methods for MCI versus NC classification

Classification methods	ACC (%)	SEN (%)	SPE (%)	Area under ROC (%)
Single classifier	79.4	79.2	79.5	87.8
Direct fusion of low-level classifiers	83.2	81.8	84.7	89.5
Proposed hierarchical fusion	**85.3**	**82.3**	**88.2**	**91.0**

TABLE VI

Comparison of classification accuracies for AD versus NC

Methods	Subjects	Features	Classifier	ACC (%)	SEN (%)	SPE (%)
(Hinrichs et al., 2009)	183 (AD + NC)	Voxel-wise GM	LP boosting	82.0	85.0	80.0
(Cuingnet et al., 2011)	137AD + 162NC	Voxel-wise GM	SVM	88.6	81.0	95.0
(Zhang et al., 2011)	51AD + 52NC	93 ROI GM volume	SVM	86.2	86.0	86.3
(Wolz et al., 2011)	198AD + 231NC	Four types of MRI features[a]	LDA	89.0	85.0	93.0
Proposed method	198AD + 229NC	Voxel-wise GM	Hierarchical fusion	**92.0**	**90.9**	**93.0**

[a]Tour types of MRI features include hippocampal volume, tensor-based morphometry, cortical thickness, and manifold-learning based features.

TABLE VII

Comparison of classification accuracies for MCI versus NC

Methods	Subjects	Features	Classifier	ACC (%)	SEN (%)	SPE (%)
Cuingnet et al. [2011]	76MCI + 162NC	Voxel-wise GM	SVM	81.17	73.00	85.00
Zhang et al. [2011]	99MCI + 52NC	93 ROI GM volume	SVM	72.00	78.5	59.6
Wolz et al. [2011]	167pMCI + 231NC	Four types of MRI features[a]	LDA	84.0	82.0	86.0
Proposed method	225MCI + 229NC	Voxel-wise GM	Hierarchical fusion	**85.3**	**82.3**	**88.2**

[a]Four types of MRI features include hippocampal volume, tensor-based morphometry, cortical thickness, and manifold-learning based features.