

# Machine Learning Techniques for AD/MCI Diagnosis and Prognosis

Chong-Yaw Wee\*, Daoqiang Zhang\*, Luping Zhou\*†, Pew-Thian Yap, Dinggang Shen

**Abstract** In the past two decades, machine learning tools have been extensively applied for the detection of neurologic or neuropsychiatric disorders, especially Alzheimer's disease (AD) and its prodrome, mild cognitive impairment (MCI). This chapter presents some latest developments in application of machine learning tools to AD and MCI diagnosis and prognosis. We will divide our discussion into single modality and multimodality approaches. We will discuss how various biomarkers as well as connectivity networks can be extracted from the individual modalities, i.e., structural T1-weighted imaging, diffusion-tensor imaging (DTI) and functional magnetic resonance imaging (fMRI), for effective diagnosis and prognosis. We will further demonstrate how these modalities can be fused for further performance improvement.

## 1.1 Background

Alzheimer's disease (AD) is the most common form dementia, characterized by cognitive and intellectual deficits that is serious enough to interfere daily life, without effective treatment. It gets worse over time by gradually destroying brain cells, causing loss in memory and the abilities to reason, make judgements and communicate. It is reported that the number of affected people is expected to double in the next 20 years, and 1 in 85 people will be affected by 2050 (Brookmeyer et al. 2007). This becomes worse as life expectancy increases. With the aging of the world population, AD has become a serious problem and a huge burden to the healthcare system, especially in developed countries. Recognizing the urgent need to slow down or completely prevent the occurrence of a worldwide healthcare crisis, effort has been under way to administer and to develop effective pharmacological and behavioral interventions for delaying the onset and progression of the disease.

---

Chong-Yaw Wee, Daoqiang Zhang, Pew-Thian Yap and Dinggang Shen  
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599,  
U.S.A., e-mail: {cywee, dqzhang, pew\_yap, dgsheng}@med.unc.edu

Luping Zhou  
Australian eHealth Center, CSIRO, Australia. e-mail: luping.zhou@csiro.au

\*Chong-Yaw Wee, Daoqiang Zhang and Luping Zhou contributed equally to this book chapter.

† The work was performed when Luping Zhou was with the Department of Radiology and BRIC, the University of North Carolina at Chapel Hill.

A significant body of literature (Johnson et al. 2006; Thompson and Apostolova 2007; Whitwell et al. 2007) suggests that pathological manifestation of AD begin many years before it can be diagnosed using cognitive tests. At the stage where symptoms can be observed, significant neurodegeneration has already occurred. Studies suggest that individuals with mild cognitive impairment (MCI), a prodrome of AD, tend to progress to probable AD at an annual rate of 10% to 15% (Grundman et al. 2004), compared with healthy controls who develop dementia at a rate of 1% to 2% (Bischkopf et al. 2002). Compared with AD, MCI is more difficult to diagnose due to the subtlety of cognitive impairment, especially in high functioning individuals who are able to maintain a positive public or professional profile without showing obvious cognitive impairment. At the present time, AD-related neurodegeneration such as structural atrophy (Jack Jr. et al. 2005), pathological amyloid depositions (Jack Jr. et al. 2010) and metabolic alterations (Nestor et al. 2004) have been identified as potential biomarkers.

Advanced statistical machine learning and pattern recognition techniques have been actively applied to map neurodegenerative patterns during the early stage of the disease where only very mild symptoms are evident (Davatzikos et al. 2010; Davatzikos et al. 2008; Fan et al. 2008a; Vemuri et al. 2009). Examples of machine learning tools that are widely used in medical imaging analysis include support vector machine (SVM) (Vapnik 1999), boosting (Morra et al. 2010), artificial neural networks (Jiang et al. 2010),  $k$ -nearest neighbor (Fitzpatrick and Sonka 2000) and linear discriminant analysis (Bankman 2008). In addition to determining group differences, pattern classification methods can be trained to identify individuals who are at risk for AD (Kloppel et al. 2008; Fan et al. 2008b; Davatzikos et al. 2008; Fan et al. 2007; Fan et al. 2008a; Fan et al. 2008c; Vemuri et al. 2008). A recent study demonstrated that classification methods are capable of identifying AD patients via their MRI scans with accuracy comparable to experienced neuroradiologists (Kloppel et al. 2008). Efforts have also been undertaken to develop regression techniques for relating clinical scores to imaging data (Duchesne et al. 2005; Chu et al. 2007; Fan et al. 2009) facilitating continuous monitoring of AD progression. In this chapter, we will focus on machine learning based diagnosis and prognosis of AD/MCI using single and multiple modality of information.

### 1.1.1 Single-Modality-Based Diagnosis and Prognosis

Single-modality-based methods rely on simpler scanning protocols and hence require less image acquisition effort, making them more clinically feasible. For example, many methods use only the structural MRI brain images for classification between AD/MCI patients and normal controls (Westman et al. 2010; Fan et al. 2008a; Lao et al. 2004). Neuroimaging measurements that are used include: re-

gional brain volumes (Chetelat and Baron 2003; Jack Jr. et al. 1998), cortical thickness (Thompson et al. 2004; Dickerson et al. 2009; Thompson et al. 2001) and hippocampal volume and shape (Chupin et al. 2009; Colliot et al. 2008).

The understanding of brain anatomical circuitry has been experiencing considerable progress due to the development of diffusion tensor imaging (DTI), which is capable of delineating white matter (WM) fiber bundles through characterization of the water diffusion (Gong et al. 2009). WM tracts connecting brain regions can be reconstructed *in vivo* using diffusion tractography (also called fiber tracking) to model the brain connectivity network (Gong et al. 2009). Diffusion measures such as fractional anisotropy (FA) and mean diffusivity (MD) are widely used in statistical analysis to localize WM changes related to AD and MCI (Rose et al. 2007; Zhang et al. 2007).

Functional connectivity is defined as the temporal correlation between regional neurophysiological signal fluctuations (Friston et al. 1993; Greicius 2008). The neurophysiological index used in functional magnetic resonance imaging (fMRI) is the blood oxygenation level dependent (BOLD) signal. BOLD signal exhibits low-frequency spontaneous fluctuations in the resting brain and shows a high degree of temporal correlation across different brain regions. Since the seminal work of Biswal et al. (Biswal et al. 1995), resting-state fMRI has been widely applied to the analysis of various neuropsychological diseases including MCI (Sorg et al. 2007) and AD (Greicius et al. 2004). One apparent advantage of resting-state fMRI over task-activation fMRI is that no complicated experimental design is required. Experiments can be performed easily by patients who may have difficulties performing specific task inside the scanner, especially those with disorders exhibiting prominent cognitive degeneration, such as AD (Greicius 2008).

Another important imaging modality for AD/MCI detection is fluorodeoxyglucose positron emission tomography (FDG-PET) (Chetelat and Baron 2003). With FDG-PET, some recent studies have reported reduction of glucose metabolism in the parietal, posterior cingulate and temporal brain regions of AD patients (Diehl et al. 2004). Besides neuroimaging techniques, biological or genetic biomarkers are effective alternatives for AD/MCI diagnosis. Researchers found that 1) the increase of cerebrospinal fluid (CSF) total tau (*t*-tau) and tau hyperphosphorylated at threonine 181 (*p*-tau) are related to neurofibrillary tangle, 2) the decrease of amyloid  $\beta$  ( $A\beta_{42}$ ) indicates amyloid plaque deposit, and 3) the presence of the apolipoprotein E (APOE)  $\epsilon 4$  allele can predict cognitive decline or conversion to AD (Fjell et al. 2010).

### 1.1.2 Multimodality-Based Diagnosis and Prognosis

Different imaging modalities provide complementary information that can be useful for AD/MCI diagnosis (Fjell et al. 2010; Landau et al. 2010; Walhovd et al. 2010b). For example, it was reported that FDG-PET and MRI measures may be complementary and differentially sensitive to memory in health and disease, with

metabolism being the stronger predictor in normal controls, and morphometry most related to memory function in AD (Walhovd et al. 2010b). Also, it is shown that morphometric changes in AD and MCI are related to CSF biomarkers, but can also provide information that is complementary to CSF biomarkers (Fjell et al. 2010). A more recent study compared the respective prognostic ability of genetic, CSF, neuroimaging, and cognitive measures obtained in the same participants, and demonstrated that complementary information provided by these different modalities can be used for enhanced AD/MCI diagnosis (Landau et al. 2010). Inspired by these findings, a number of studies used two or more biomarkers simultaneously for detection of AD and MCI: MRI and CSF in (Davatzikos et al. 2010; Vemuri et al. 2009), MRI and cognitive testing in (Geroldi et al. 2006), MRI and PET in (Fan et al. 2008c; Hinrichs et al. 2009a), MRI and APOE biomarkers in (Ye et al. 2008), FDG-PET and CSF in (Fellgiebel et al. 2007), FDG-PET and cognitive testing in (Chetelat et al. 2005), and MRI, CSF, and FDG-PET in (Walhovd et al. 2010a).

## 1.2 Single-Modality-Based Diagnosis and Prognosis

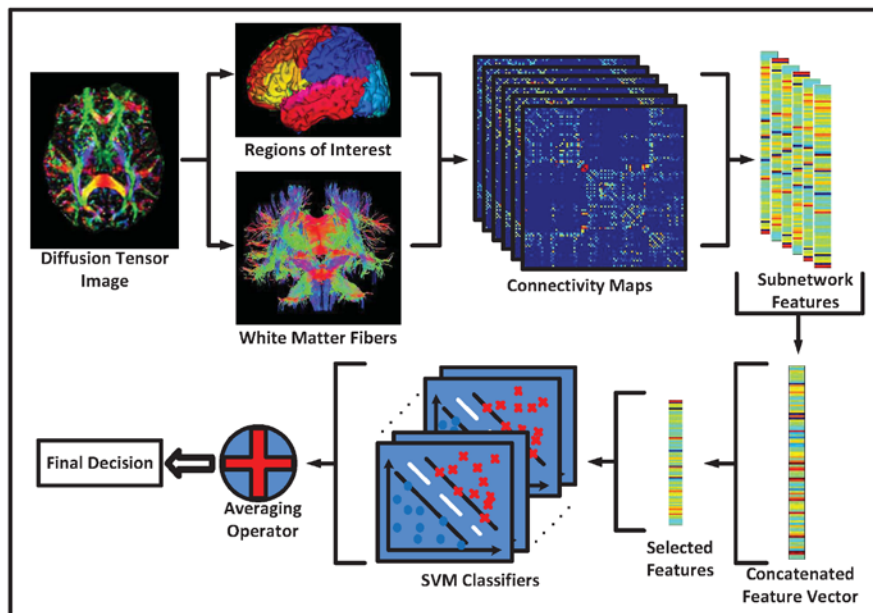
Progressive degenerative neurological diseases such as AD and similar dementias exhibit spatially and temporally diffuse pathology, where the brain is damaged in a large-scale, highly connected network, rather than in one single isolated region. In view of this, designing a description of interregional connections that is more sensitive to the pathology is necessary for accurate diagnosis of the disease. In this section, we will discuss some recently proposed descriptions based on single modality biomarkers for AD and MCI diagnosis and prognosis. Models of whole-brain connectivity, which comprise networks of brain regions connected either by anatomical tracts or functional associations, have drawn a great deal of interest recently due to the increasing reliability of network characterization through neurobiologically meaningful and computationally efficient measures (Hagmann et al. 2008; Sporns and Zwi 2004).

### 1.2.1 Structural Analysis via Enriched White Matter Connectivity Networks

Recently, an enriched description of WM connections via diffusion tractography (Wee et al. 2011a) was proposed to convey rich information related to the topological and biophysical properties of the connection. This is achieved by using a collection of physiological parameters that are derived during whole-brain streamline fiber tractography to effectively describe small variations on WM regions caused by pathological attacks. This is in contrast to a simple connectivity descrip-

tion using a single physiological parameter that affords only limited information. The proposed MCI classification framework is shown graphically in Fig. 1.

Physiological parameters which are included in the enriched description are fiber count, functional anisotropy (FA), mean diffusivity (MD), and principal diffusivities ( $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ ). During tractography, the number of fibers passing through each pair of regions is counted. Two regions are considered as anatomically connected if fibers passing through their respective masks are present. Considering the connection between every possible pair of regions gives us the connection topology of the network. Connectivity networks of FA, MD and principal diffusivity are derived at the same time by taking their average values along the connecting fibers. These five connectivity networks share identical connection topology as defined by the fiber count network, but convey different biophysical properties. Dimensional of the constructed connectivity networks is depending on the spatial scale of anatomical parcellation atlas used. An example of the six connectivity networks for one subject is provided in Fig. 2.

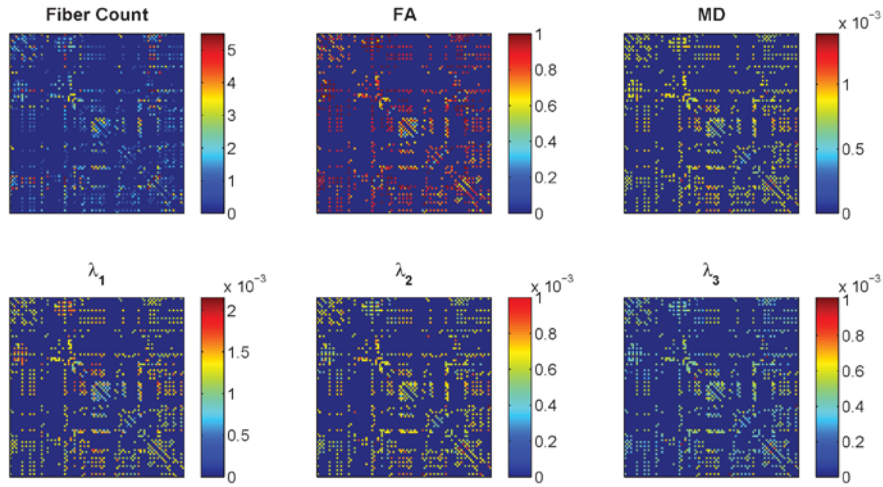


**Fig. 1** Classification based on enriched description of WM connections.

Measures in complex network analysis typically quantify connectivity profiles associated with the nodes and reflect the way in which these nodes are embedded in the network. Clustering coefficient (Rubinov and Sporns 2010; Watts and Strogatz 1998), which quantifies the degree to which nodes in a network tend to cluster together, is widely used to extract information from the constructed brain connectivity networks for group analysis. To prevent the smooth out of subtle variations caused by pathological effects on the constructed networks, the

weighted local clustering coefficient instead of its global version is computed from connectivity networks.

In neuroimaging, it is normal to obtain a relatively high dimensional feature pool and a good classification performance is difficult to achieve if all the extracted features are directly used. This is due to the fact that some features are less sensitive, irrelevant or redundant for classification, compared to others. By using a proper feature selection procedure, an optimal subset of features, which possesses the most discriminative power, is selected to improve generalization performance.



**Fig. 2** Connectivity networks constructed with different physiological parameters.

The discriminative power of a feature can be qualitatively measured by its relevance to classification as well as its generalizability. The relevance of a feature to classification can be measured through its correlation with clinical labels (Fan et al. 2007). Pearson correlation coefficient, a widely used linear correlation measure in machine learning, is used to rank features. The larger the absolute value of the Pearson correlation coefficient, the more relevant the feature is to classification.

The generalizability of a feature is evaluated via leave-one-out cross-validation when measuring the correlation of the feature with respect to the clinical labels via Pearson correlation coefficient (Fan et al. 2007). For  $n$  training samples, the worst absolute Pearson correlation coefficient resulting from  $n$  leave-one-out correlation measurement is conservatively selected as the effective correlation coefficient. This approach is particularly important when examining a very large number of features to minimize the effect of outliers.

Nevertheless, the correlation or ranking score is computed independently for each feature, without considering the correlation with other features. This inevitably causes some redundant features to be selected and eventually affects the classification performance. In order to minimize this effect, a wrapper-based feature se-

lection method called SVM-RFE algorithm (Guyon et al. 2004; Rakotomamonjy 2003) is used to select the final optimal subset based on the feature ranking. In this algorithm, SVM is used as classifier to evaluate the discriminative power of the selected subset of features.

In (Wee et al. 2011a), the classification performance of the enriched WM connectivity description method is evaluated using a nested full leave-one-out cross-validation strategy to ensure a relatively unbiased estimate of the generalization power of the classifiers to new subjects. In each leave-one-out case, one subject is first left out as the testing subject, and the remaining subjects are used for feature extraction, feature selection and classifier training. Then, second or inner leave-one-out is applied within the training set, to build an ensemble classifier whose parameters are automatically optimized. Specifically, for  $n$  total number of subjects involved in the study, one is left out for testing, and the remaining  $n-1$  are used for training. From these  $(n-1)$  samples,  $(n-1)$  different training subsets are formed by each time leaving one more sample out, giving us  $(n-2)$  subjects in each training subset. For each subset, a SVM classifier is built with its performance evaluated using the second left out subject. This procedure is repeated for  $n-1$  times, once for each training subset. This procedure allows us to select parameters which maximize the area under receiver operating characteristic (ROC) curve. When the completely unseen (totally left out during the entire training and parameter optimization process) test sample is to be classified, all  $(n-1)$  classifiers are used, and their outcomes are combined using an averaging operator to provide the final classification decision. This process is repeated  $n$  times, each time leaving out a different subject, finally leading to a overall cross-validation classification accuracy.

### 1.2.2 Functional Analysis via Multi-Spectral Connectivity Networks

Over the past several years, resting-state fMRI (rs-fMRI) has emerged as a novel informative method for investigating the development of large-scale functional networks in the human brain. This method, first used to demonstrate coherent spontaneous low-frequency fluctuations in BOLD signal within the adult somatomotor system (Biswal et al. 1995), involves measuring the hemodynamic response related to neural activity in the brain or spinal cord from participants as they lay in the MRI scanner in the “resting state”. It is recently employed to identify individuals with MCI from normal controls and achieved a reasonably well performance (Wee et al. 2011b).

In (Wee et al. 2011b), Wee et al. suggested an efficient characterization of rs-fMRI time series via: 1) Gray matter (GM) masking to elucidate blood oxygenation level dependent (BOLD) signal variations in GM region by factoring out the contribution of WM and cerebrospinal fluid (CSF); 2) Multi-spectral characterization to quantify relatively small changes of BOLD signal by decomposing the

mean time series of each ROI into five distinct frequency sub-bands; and 3) Graph theoretic analysis to characterize topological properties and strengths of brain functional connectivity networks through neurobiologically meaningful and computationally efficient measures (Sporns and Zwi 2004; Bassett and Bullmore 2006; Hagmann et al. 2008).

*In vivo* neuroimaging studies suggest that normal aging (Courchesne et al. 2000) as well as AD (Karas et al. 2003; Thompson et al. 2003) are associated with GM volume loss. There is an emerging body of evidence that MRI can observe deterioration, including progressive loss of GM in the brain, from MCI to full-blown AD (Whitwell et al. 2008). It has been shown that the GM volume of human brain decreases linearly by approximately 5% per decade throughout lifetime after 9 years of age (Courchesne et al. 2000). It has been reported that local GM loss rates are approximately 5.3% and 0.9% per annum in AD and healthy aging respectively, with an asymmetric trend where faster loss rate is observed in the left hemisphere than in the right (Thompson et al. 2003). Furthermore, removal of signal from the ventricles and WM can reduce the relatively high proportion of noise caused by the cardiac and respiratory cycles (Van Dijk et al. 2010). Accordingly, tissue segmentation is performed on T1-weighted image of each subject to obtain the GM, WM and CSF. Then, the segmented GM image is used to mask the fMRI images. This procedure eliminates the possible contribution from WM and CSF in the fMRI time series. Anatomical parcellation is further used to divide the brain into regions-of-interest (ROIs).

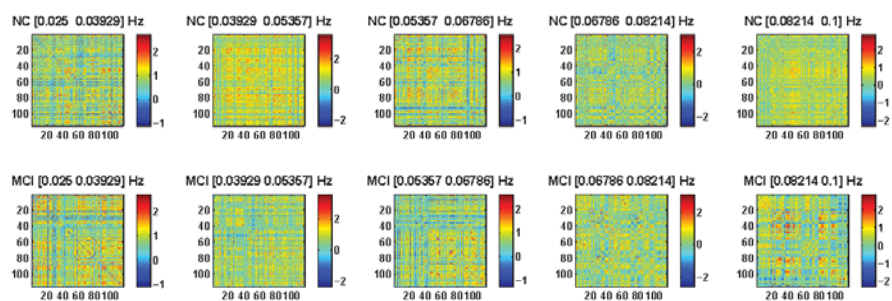
For each subject, the mean time series of each individual ROI is obtained by averaging the GM-masked fMRI time series over all voxels in that particular ROI. Temporal band-pass filtering with frequency interval ( $0.025 \leq f \leq 0.100\text{Hz}$ ) is then performed on the mean time series of each individual ROI. This frequency interval provides a reasonable trade-off between avoiding the physiological noise associated with higher frequency oscillations (Cordes et al. 2001) and the measurement error associated with estimating very low frequency correlations from limited time series (Achard et al. 2008). In conventional approach, the regional mean time series of whole spectrum is directly employed to construct functional connectivity networks. However, this whole spectrum description, a relatively “global” analysis, might not be sensitive enough to delineate complex yet subtle pathological patterns related to the neurological disease.

In (Wee et al. 2011b), a multi-spectral characterization of the regional mean time series, which utilizes multiple frequency sub-bands, is proposed to construct functional connectivity networks. The band-pass filtered GM-masked mean time series of each region is decomposed into five distinct, equally divided frequency sub-bands using the Fast Fourier transform (FFT). By using this multi-spectral characterization, a relatively “local” analysis which preserves most of the small variations on BOLD signal, can be achieved.

Functional connectivity that examines interregional correlations in neuronal variability (Friston et al. 1993) can be measured using a pairwise Pearson correlation coefficients between a given pair of ROIs. Given a set of  $N$  random variables, the

Pearson correlation matrix is a symmetric matrix in which each off-diagonal element is the correlation coefficient between a pair of variables. The brain regions can be considered as a set of nodes and the correlation coefficients as signed weights on the set of edges. A Fisher's  $r$ -to- $z$  transformation is applied on the computed Pearson correlation matrix to improve the normality of Pearson correlation coefficients. The feature extraction, feature selection and high-dimensional multivariate classification used for MCI diagnosis in (Wee et al. 2011a) are applied in the case of rs-fMRI.

Examples of the functional connectivity maps constructed using the proposed multi-spectral characterization for one normal control (NC) and one MCI patient are shown in Fig. 3.



**Fig. 3** Multispectral functional connectivity maps for a normal control (NC) and an MCI patient.

### 1.2.3 Hierarchical Brain Networks from T1-Weighted MRI

Owing to its clinical accessibility, T1-weighted MRI has been extensively studied for the diagnosis and prognosis of MCI and AD. Conventionally, the mean tissue volumes of GM, WM and CSF are usually calculated locally within ROI, and used as features for classification. Nevertheless, disease-induced brain structural changes may not happen at isolated spots, but in several inter-related regions. To deal with this problem, it is proposed in (Zhou et al. 2011) that representing the brain as a system of interconnected regions is a more effective way of characterizing subtle changes than by using local isolated measures. For this purpose, a hierarchical brain network is constructed to directly model the pairwise ROI interactions within a subject as features for classification, with each node denoting an ROI and each edge characterizing the pairwise connection. In particular, each ROI is characterized by a volumetric vector that consists of the volumetric ratios of GM, WM and CSF in this ROI. The interaction between two ROIs within the same subject is computed as the Pearson correlation of the corresponding volumetric elements. The correlation value measures the similarity of the tissue compositions between a pair of brain regions. When a patient is affected by MCI, the cor-

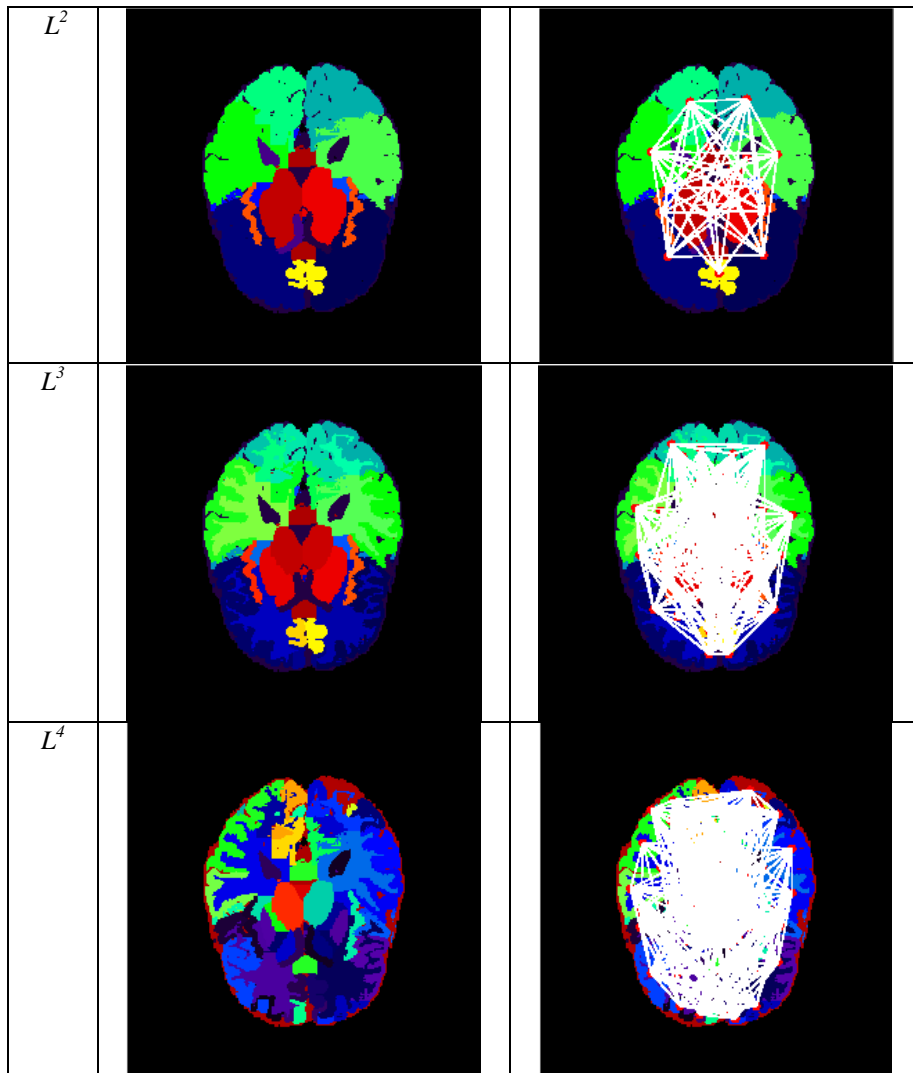
relation values of some brain regions with other regions will be affected, due possibly to the factors such as tissue atrophy.

By computing the pairwise correlation between ROIs, the approach in (Zhou et al. 2011) provides a second order measure of the ROI volume, while the conventional approaches only employ the first order measure of the volume. As higher order measures, the proposed new features may be more descriptive, but also more sensitive to noise, such as registration errors. Therefore, a four-layer hierarchy of multi-resolution ROIs (Fig. 4 (a)) is introduced to increase the robustness of classification. Effectively, the correlations are considered at different scales of regions, thus giving different levels of noise suppression and discriminant information, which can be sieved by the proposed classification scheme. This approach considers the correlations both within and between different resolution scales (Fig. 5), because a certain “optimal” scale often cannot be known a priori. The brain network may be quite complicated. For instance, Fig. 4 (b) partially shows the network connections between ROIs in each layer. To efficiently obtaining the informative network features, a membership matrix is created to indicate the relationship of ROIs from different layers. The membership matrix is computed offline: it is fixed once the hierarchical structure has been determined. For a new brain image, it only needs to compute the ROI interactions on the bottommost layer that has the finest ROIs, and then propagate the correlations to other layers effectively via this membership matrix.

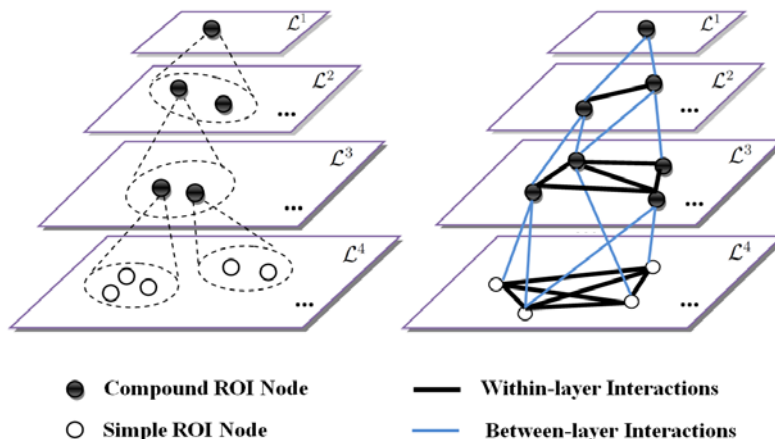
Note that the proposed brain network may not have the property of small-worldness (sparseness) as shown in DTI and fMRI networks (Bassett and Bullmore 2006), because the connections in this case are not based on functions or real neuron-connections. The dense adjacency matrix resulting from the correlation of tissue compositions implies that WM, GM and CSF fractions of many different brain regions are consistently similar. Note that the far-away region pairs can have meaningful tissue composition similarity, since distance information is not included in this approach. Because the network is fully connected, some commonly used network features, such as local clustering coefficients, do not work efficiently as they do for sparse networks in DTI and fMRI. Therefore, the weights of edges are directly used as features, that is, the elements in the upper triangle matrices of correlation matrices are concatenated to form the feature vectors.

Compared with conventional methods, this approach produces a significant larger pool of features, which if improperly dealt with, will result in intractability when used for classifier training. Conventionally, there are usually two ways to deal with the high dimensionality of features: i) select a subset of the most discriminative features from the original feature set, known as feature selection, or ii) combine the original features linearly or non-linearly to get a lower dimensional feature space, known as feature embedding. In (Zhou et al. 2011), based on the characteristics of the network features, a dimensionality reduction process is proposed to efficiently reduce the feature dimensionality to a manageable level while at the same time preserving discriminative information as much as possible. It couples both feature selection and feature embedding via Partial Least Square

(PLS) analysis (Rosipal and Kramer 2006) in an integrated optimization process. As a supervised learning method, PLS considers about the information in the classification labels and thus achieves a better discrimination than many of the commonly used unsupervised methods, for example, Principal Components Analysis (PCA) and the Laplacian Eigenmap, and even some commonly used supervised methods, such as Kernel Fisher Discriminant Analysis (KFDA).



**Fig. 4** (a) Hierarchical ROIs in three different layers (the top layer is a whole brain which is not shown); (b) Network connections between ROIs within different layers.



**Fig. 5** Explanation of the network model. **Left:** Two types of nodes are included in the hierarchical network: the simple node in the bottommost layer, and the compound node in other layers. Each compound node is obtained by grouping several simple nodes agglomeratively. **Right:** Two types of edges are included in the hierarchical network, each modeling the within-layer and between-layer interactions, respectively.

Taking advantages of PLS analysis, the approach in (Zhou et al. 2011) achieves good classification and generalization in four steps:

*In Step 1*, the discriminative power of a feature is measured by its relevance to classification. The relevance is computed by the Pearson correlation between each original feature and the classification label. The larger the absolute value of the correlation, the more discriminative the feature. Features with correlation values lower than a threshold are filtered out.

*In Step 2*, a subset of features is further selected from the result of Step 1 in order to optimize the performance of PLS embedding in Step 3. In particular, a PLS model is trained using the selected features from Step 1. Then a method called Variable Importance on Projection (VIP) (Wold et al. 1993) is used to rank these features according to their discriminative power in the learned PLS model. The discriminative power is measured by a VIP score. The higher the score, the more discriminative the feature. About 60 ~ 80 features with the top VIP scores are selected for feature embedding in the next step.

*In Step 3*, using the features selected in Step 2, a new PLS model is trained to find an embedding space which best preserves the discrimination of features. The embedding is performed by projecting the feature vectors in the original feature matrix onto the new weight vectors learned by PLS analysis.

*In Step 4*, after PLS embedding, a small number of features in the new space are able to capture the majority of the class discrimination. This greatly reduces the complexity of relationships between data. Therefore, these features are used to train a linear SVM for predicting MCI patients and normal controls.

Note that the number of selected features in each step is determined by cross-validation on the training data.

The merits of the proposed method are recapped as follows. Firstly, the proposed method utilizes a second-order volumetric measure that is more descriptive than the conventional first-order volumetric measure. Secondly, while the conventional approaches only consider local volume changes, the proposed method considers global information by pairing ROIs that may be spatially far away. Thirdly, the proposed method seamlessly incorporates both the local volume features and the proposed global network features into the classification by introducing a whole-brain ROI at the top of the hierarchy. By correlating with the whole-brain ROI, each ROI can provide a first order measurement of local volume. Fourthly, the proposed method involves only linear methods, leading to easy interpretations of the classification results. Note that the interpretation is equally important as classification in neuroimaging analysis. Finally, for the first time, the proposed method investigates the *relative* disease progression speeds in different regions, providing a complementary perspective of the spatial atrophy patterns to conventional methods.

### 1.3 Multimodality-Based Diagnosis and Prognosis

A number of studies have shown that biomarkers from different modalities may contain complementary information for diagnosis of AD/MCI, and several works on combining different modalities have been reported (Walhovd et al. 2010b; Fellgiebel et al. 2007; Davatzikos et al. 2010; Fan et al. 2008c; Hinrichs et al. 2009b; Vemuri et al. 2009; Ye et al. 2008). A common practice in these works is to concatenate all features (from different modalities) into a long feature vector. However, approaches as such do not distinguish between the different modalities and are hence not the best way of combining information from multiple sources. In this section, we provide an alternative way by using multiple kernel combination to integrate different biomarkers. Compared with the direct feature concatenation method, the kernel combination method has the following advantages: 1) it provides a unified way to combine heterogeneous data when different types of data cannot be directly concatenated; 2) it offers more flexibility by using different weights on biomarkers of different modalities. To overcome the small sample size problem in training multimodality classifier, we adopt a semi-supervised learning technique that can learn from both labeled and unlabeled data.

#### 1.3.1 Multimodality Data Fusion via Multi-Kernel SVM

A general framework based on kernel methods (Scholkopf and Smola 2002) was recently proposed by Zhang et al. in (Zhang et al. 2011) to combine multiple biomarkers (i.e., MRI, PET and CSF) for discriminating between AD (or MCI) and normal controls. The proposed method is based on kernel combination and can be easily embedded into the conventional SVM classifier for high-dimensional

pattern classification. Moreover, unlike other kernel combination methods which can only process one data type, i.e., numbers, this method is capable of combining numeric data, strings, and graphs. The framework proposed by Zhang et al. (Zhang et al. 2011) is detailed in the following.

In SVM, linearly non-separable samples are mapped, through a kernel-induced implicit mapping function, to a higher or even infinite dimensional space, where they are more likely to be linearly separable than in the original lower-dimensional space. A maximum margin hyperplane is then sought in the higher-dimensional space. Multiple-kernel learning (MKL), which is pioneered by Lanckriet et al. (Lanckriet et al. 2004b) and Bach et al. (Bach et al. 2004), is an additive extension of the single kernel SVM by incorporating multiple kernels. Suppose that we are given  $n$  training samples and each of them is of  $M$  modalities. Let  $\mathbf{X}_i^{(m)}$  denote a feature vector of the  $m$ -th modality of the  $i$ -th sample, and its corresponding class label be  $y_i \in \{-1, 1\}$ . Multiple-kernel based SVM solves the following primal problem:

$$\begin{aligned} \min_{\mathbf{w}_m^{(m)}, b, \xi_i} \quad & \frac{1}{2} \sum_{m=1}^M \beta_m \|\mathbf{w}^{(m)}\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i \left( \sum_{m=1}^M \beta_m \left( (\mathbf{w}^{(m)})^T \phi^{(m)}(\mathbf{x}_i^{(m)}) + b \right) \right) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (1.1)$$

where  $\mathbf{W}^{(m)}$ ,  $\phi^{(m)}$  and  $\beta_m \geq 0$  denote the normal vector of hyperplane, the kernel-induced mapping function and the weighting factor of the  $m$ -th modality, respectively.

Similar to the conventional SVM, the dual form of multiple-kernel SVM can be represented as below:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{m=1}^M \beta_m k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, n. \end{aligned} \quad (1.2)$$

where  $k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) = \phi^{(m)}(\mathbf{x}_i^{(m)})^T \phi^{(m)}(\mathbf{x}_j^{(m)})$  is the kernel function for the two training samples on the  $m$ -th modality.

For a new test sample  $\mathbf{x} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(M)}\}$ , we first denote  $k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}^{(m)}) = \phi^{(m)}(\mathbf{x}_i^{(m)})^T \phi^{(m)}(\mathbf{x}^{(m)})$  as the kernel between the new test sample

and each training sample on the  $m$ -th modality. Then, the decision function for the predicted label can be obtained as below:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^n y_i \alpha_i \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x^{(m)}) + 1 \right) \quad (1.3)$$

Multiple-kernel based SVM can be naturally embedded into the conventional single-kernel SVM if we interpret  $k(\mathbf{x}_i, \mathbf{x}_j) = \sum_m \beta_m k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)})$  as a mixed kernel between the multimodality training samples  $\mathbf{X}_i$  and  $\mathbf{X}_j$ , and  $k(\mathbf{x}_i, \mathbf{x}) = \sum_m \beta_m k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}^{(m)})$  as a mixed kernel between the multimodality training sample  $\mathbf{X}_i$  and the test sample  $\mathbf{X}$ . In fact, this method collapses multiple kernels into one.

It is worth noting that the multiple-kernel SVM proposed by Zhang et al. in (Zhang et al. 2011) is different from the existing multi-kernel learning methods (Wang et al. 2008; Lanckriet et al. 2004a). One key difference is that the weights  $\beta_m$ s are not jointly optimized with other SVM parameters (e.g.,  $\alpha$ ). Instead, Zhang et al. enforce the constrain  $\sum_m \beta_m = 1$  and use a coarse-grid search through cross-validation on the training samples to determine the optimal values. The obtained  $\beta_m$  values are used to combine multiple kernels into a mixed kernel, which can be incorporated into the standard SVM to be solved using conventional SVM solvers, e.g., LIBSVM (Chang and Lin 2001).

### 1.3.2 Semi-Supervised Learning Using Unlabeled Data

One challenge in AD patient identification is that the number of AD patients and normal controls (NCs) is generally very small, thus making it difficult to train an effective classifier. As a remedy, we note that MCI subjects, although their cognitive status is uncertain, can be helpful for improving classifier construction. To exploit the potential of using MCI subjects to aid classification between AD and NC subjects, Zhang et al. (Zhang and Shen 2011) treat MCI subjects as unlabeled data (i.e., not classified either as AD or NC), and then employ a semi-supervised learning technique (Tiwari et al. 2010; Chapelle et al. 2006) to solve the classification problem. In the following, we will first introduce the semi-supervised learning technique, called Laplacian Regularized Least Squares (LapRLS) method (Belkin et al. 2006), and then derive its multimodality extension (mLapRLS).

### 1.3.2.1 Laplacian Regularized Least Squares (LapRLS)

Assume we have  $l$  labeled data (from AD and NC samples),  $(x_i, y_i)$ ,  $i=1, \dots, l$ , and  $u$  unlabeled data (from MCI samples),  $(x_j, y_j)$ ,  $j=l+1, \dots, l+u$ . Suppose  $k(\cdot, \cdot)$  is a Mercer kernel function, and let  $H$  be the associated Reproducing Kernel Hilbert Space (RKHS) and  $\|\cdot\|$  be the corresponding norm. The LapRLS algorithm solves the following least-squared loss function (Belkin et al. 2006):

$$\min_{f \in H} \frac{1}{l} \sum_{i=1}^l (y_i - f(x_i))^2 + \gamma_A \|f\|^2 + \frac{\gamma_B}{(u+l)^2} \mathbf{f}^T \mathbf{L} \mathbf{f} \quad (1.4)$$

where  $\mathbf{f} = [f(x_1), \dots, f(x_{l+u})]^T$ .  $L$  is the graph Laplacian given as  $L=D-W$ , where  $W_{ij}$  are the edge weights in the adjacency graph defined on both labeled and unlabeled data and the diagonal matrix  $\mathbf{D}$  is given by  $D_{ii} = \sum_j W_{ij}$ . Symbols  $\gamma_A$  and  $\gamma_B$  are the two regularization parameters. Intuitively, the first two terms in Eqn. (4) are for the supervised learning on only the labeled data (AD and NC samples), while the last term in Eqn. (1.4) involves both labeled and unlabeled data (AD, NC and MCI samples) for unsupervised estimation of the intrinsic geometric structure of the whole data. According to the Representer Theorem (Belkin et al. 2006), the solution to Eqn. (1.4) is an expansion of kernel functions over both labeled and unlabeled data:

$$f(x) = \sum_{i=1}^{l+u} \alpha_i k(x, x_i) \quad (1.5)$$

Substituting Eqn. (1.5) into Eqn. (1.4), we arrive at the dual form of Eqn. (1.4) with respect to the  $(l+u)$ -dimensional variable vector  $\alpha = [\alpha_1, \dots, \alpha_{l+u}]^T$ :

$$\min_{\alpha \in \mathbb{R}^{l+u}} \frac{1}{l} (Y - JK\alpha)^T (Y - JK\alpha) + \gamma_A \alpha^T K \alpha + \frac{\gamma_B}{(u+l)^2} \alpha^T K L K \alpha \quad (1.6)$$

where  $K = \{k(x_i, x_j)\}$  is an  $(l+u) \times (l+u)$  kernel matrix over all labeled and unlabeled data;  $Y = [y_1, \dots, y_l, 0, \dots, 0]$  is an  $(l+u)$ -dimensional label vector, and  $J = \text{diag}(1, \dots, 1, 0, \dots, 0)$  is an  $(l+u) \times (l+u)$  diagonal matrix with the first  $l$  diagonal entries as 1 and the rest as 0. By computing the derivative of Eqn. (1.6) with respect to  $\alpha$ , and let it be zero, we obtain the following solution:

$$\alpha = \left( JK + \gamma_A \mathbf{I} + \frac{\gamma_B l}{(u+l)^2} \right)^{-1} Y \quad (1.7)$$

where  $\mathbf{I}$  is the identity matrix. It is worth noting that, when  $\gamma_B$ , Eqn. (1.7) gives zero coefficients over the unlabeled data, and the coefficients over the labeled data are exactly those given by the standard Regularized Least Squares (RLS) method, i.e., LapRLS degenerates to RLS.

### 1.3.2.2 Multimodality LapRLS (mLapRLS)

Now, we derive the multimodality extension of LapRLS, called mLapRLS, for classification between AD and NC. Given  $l$  labeled data (from AD and NC samples),  $(x_i, y_i)$ ,  $i=1, \dots, l$ , and  $u$  unlabeled data (from MCI samples),  $(x_j, y_j)$ ,  $j=l+1, \dots, l+u$ , we assume each data  $x_i$  is now composed of  $M$  modalities, i.e.,  $x_j = \{x_j^{(1)}, \dots, x_j^{(M)}\}$ ,  $i=1, \dots, l+u$ .

Define the distance function between two multimodality data  $x_i$  and  $x_j$  as

$$d(x_i, x_j) = \sum_{m=1}^M \beta_m d^{(m)}(x_i^{(m)}, x_j^{(m)}) \quad (1.8)$$

where  $d^{(m)}(\cdot, \cdot)$  denotes the distance function on the  $m$ -th modality, and  $\beta_m$ s are the nonnegative weighting parameters used to balance the contributions of different modalities. All  $\beta_m$ s are constrained by  $\sum_m \beta_m = 1$ . According to Eqn. (1.8), we can compute the adjacency graph for the multimodality data, and then obtain the corresponding edge weights matrix  $W$  and graph Laplacian  $L$  on the multimodality data. Next, we can define the kernel function on two multimodality data  $x$  and  $x_i$  as

$$k(x, x_i) = \sum_{m=1}^M \beta_m k^{(m)}(x^{(m)}, x_i^{(m)}) \quad (1.9)$$

where  $k^{(m)}$  denotes the kernel matrix over the  $m$ -th modality, similar to the definition given above for the single modality case. With the definition of  $k(x, x_i)$ , the  $(l+u) \times (l+u)$  kernel matrix  $K$  on the multimodality data can be straightforwardly obtained as  $K = k(x_i, x_j)$ . Once we have the graph Laplacian  $L$ , the definition of the kernel function  $k(x, x_i)$  on the multimodality data, and the kernel matrix  $K$ , the mLapRLS solution to the multimodality data can be obtained exactly the same as LapRLS in Eqn. (1.7). Similar to LapRLS, mLapRLS will degenerate to the corresponding multimodality RLS (mRLS) when  $\gamma_B = 0$ . In this case, mRLS uses only AD and NC samples for training.

## 1. 4 AD/MCI Diagnosis and Prognosis

In this section, we will evaluate the machine learning based classification techniques, discussed in the previous sections, for AD/MCI diagnosis and prognosis using single and multiple modality data.

### 1.4.1 Single-Modality-Based Diagnosis and Prognosis

#### 1.4.1.1 MCI Diagnosis using Enriched White Matter Connectivity Description

This study involved 27 participants (10 MCI patients and 17 socio-demographically matched healthy controls) who were recruited by the Duke-UNC Brain Imaging and Analysis Center, North Carolina, USA. Informed consent was obtained from all participants, and the experimental protocols were approved by the institutional ethics board. Confirmation of diagnosis for all subjects was made via expert consensus panels at the Joseph and Kathleen Bryan Alzheimer's Disease Research Center (Bryan ADRC) and the Department of Psychiatry at Duke University Medical Center. Diagnosis was based upon available data from a general neurological examination, neuropsychological assessment evaluation, collateral and subject symptom and functional capacity reports. Demographic information of the participants involved in this study is shown in Table 1.1.

**Table 1.1 Demographic information of the participant involved in the study.**

-	MCI	Normal
No. of Subjects	10	17
No. of males	5	8
Age (mean $\pm$ SD)	74.2 $\pm$ 8.6	72.1 $\pm$ 8.2
Years of education (mean $\pm$ SD)	17.7 $\pm$ 4.2	16.3 $\pm$ 2.4
MMSE score (mean $\pm$ SD)	28.4 $\pm$ 1.5	29.4 $\pm$ 0.9

*A priori* knowledge of the number of features that should be used for classification is not available and this number is automatically determined as part of inner loop of the nested leave-one-out cross-validation. Although it generally yields slightly lower classification performance, the nested leave-one-out cross-validation provides a better indicator of the generalizability of a classifier. The classification accuracy by the enriched description of WM connections (with six parameters) is 88.9%, which is at least an 14.8% increment from that using any single physiological parameter. The area under receiver operating characteristic (ROC) curve (AUC) of the enriched description method is 0.929, indicating its excellent diagnostic power. It is noteworthy that simple connectivity description, in most of the cases, is unable to provide good generalization power, as indicated by

the much smaller AUC values. The classification performance of the enriched and simple connectivity descriptions is summarized in Table 1.2.

**Table 1.2 Classification performance and AUC values for enriched and simple connectivity descriptions.**

Description	Accuracy (%)	AUC
Enriched	88.89	0.929
Fiber count	70.37	0.653
FA	74.07	0.859
MD	59.26	0.647
$\lambda_1$	59.26	0.629
$\lambda_2$	55.56	0.594
$\lambda_3$	59.26	0.612

The SVM-RFE algorithm (Guyon et al. 2004) is used to select a subset of most discriminant features in a backward sequential fashion by removing one feature at a time. This subset of features yields the best classification performance based on the training set. Each feature corresponds to a ROI, and a region, by capturing discriminative information, is indicative that it might be affected by the disease. Since the selected subset of features might be different for each leave-one-out case, the most significant ROIs are defined as the regions which are selected the most in all leave-one-out cases. The most discriminant regions that are selected for classification included 1) rectus gyrus, which is located on the orbital surface of the frontal lobe; 2) insula, which is located within lateral fissure between the temporal lobe and the frontal lobe; and 3) precuneus, which is a part of the superior parietal lobe hidden in the medial longitudinal fissure between the two cerebral hemispheres.

It is noteworthy that the proposed method is based on the assumption that the set of brain measurements that optimally differentiate between MCI patients and cognitively normal individuals cannot be known *a priori*, but can only be determined from the data. The leave-one-out cross-validation used here fundamentally guards against data overfitting, a persistent problem in high dimensionality analyses of datasets with relatively small sample size.

#### 1.4.1.2 MCI Diagnosis using Multi-Spectral Connectivity Characterization

This study involved 37 participants (12 MCI patients and 25 socio-demographically matched healthy controls) who were recruited by the Duke-UNC Brain Imaging and Analysis Center, North Carolina, USA. Informed consent was obtained from all participants, and the experimental protocols were approved by the institutional ethics board. Demographic information of the participants involved in this study is shown in Table 1.3.

**Table 1.3 Demographic information of the participant involved in the rs-fMRI study.**

-	MCI	Normal
No. of Subjects	12	25
No. of males	6	9
Age (mean $\pm$ SD)	75.0 $\pm$ 8.0	72.9 $\pm$ 7.9
Years of education (mean $\pm$ SD)	18.0 $\pm$ 4.1	15.8 $\pm$ 2.4
MMSE score (mean $\pm$ SD)	28.5 $\pm$ 1.5*	29.3 $\pm$ 1.1

\*One of the patients does not have a MMSE score.

Classification performance for the multi-spectral characterization of rs-fMRI regional mean time series was compared with the conventional whole-spectral characterization. The nested leave-one-out cross-validation training and testing procedures described in the Sec. 1.2.1 were applied to both methods during comparison.

The effectiveness of GM-masked and unmasked BOLD signals was evaluated in relation to the whole- and multi-spectral characterization methods. The comparison results are shown in Table 1.4.

**Table 1.4 Classification accuracies and AUC values of GM-masked and unmasked approaches using whole- and multi-spectral network characterization methods, respectively.**

Approach	Accuracy (%)	AUC
Unmasked + Whole-Spectrum	56.76	0.530
GM-Masked + Whole-Spectrum	59.46	0.543
Unmasked + Multi-Spectrum	67.57	0.620
GM-Masked + Multi-Spectrum	86.49	0.863

In agreement with the hypothesis, GM-masked BOLD signal with multi-spectral characterization outperforms the unmasked and whole-spectrum characterization methods. GM-masking, when used with the conventional whole-spectral characterization, only shows slightly improvement in terms of classification accuracy and AUC value. However, when combined with the multi-spectral characterization, the classification accuracy increases by more than 18.9% while the AUC value increases by more than 0.24, indicating significant improvement in diagnostic power. This marked improvement in performance demonstrates the effectiveness and robustness of the GM-masked multi-spectral characterization in providing relatively fine and localized analysis.

The most discriminant regions that are selected for classification are mainly located in prefrontal cortex areas and temporal lobes. The selected regions involved parts of frontal lobe such as rectus gyrus, orbitofrontal cortex and frontal gyrus, parts of temporal lobe such as temporal poles, amygdala and parahippocampal gyrus, superior occipital gyrus of occipital lobe and precuneus of parietal lobe.

### 1.4.1.3 MCI Diagnosis Using Hierarchical Brain Networks

In this study, 125 normal control subjects and 100 progressive MCI (P-MCI) subjects were taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([www.loni.ucla.edu /ADNI](http://www.loni.ucla.edu/ADNI)). The ADNI database contains approximately 200 cognitively normal elderly subjects to be followed for 3 years, 400 subjects with MCI to be followed for 3 years, and 200 subjects with early AD to be followed for 2 years. The P-MCI subjects are those who developed probable AD after the baseline scanning. The diagnosis of AD is made according to the NINCDS/ADRDA criteria (McKhann et al. 1984) for probable AD. The demographic and clinical information of all the selected subjects are summarized in Table 1.5.

**Table 1.5 Demographic information of the subjects involved in the study.**

-	P-MCI	Normal
No. of subjects	100	125
No. and percentage of males (%)	57 (57.0)	61 (48.8)
Baseline age (mean $\pm$ SD)	75.0 $\pm$ 7.1	76.1 $\pm$ 6.1
Baseline MMSE score (mean $\pm$ SD)	26.5 $\pm$ 1.7	29.1 $\pm$ 1.0

The effectiveness of constructing hierarchical brain network from T1-weighted MRI for MCI prediction was evaluated by the comparison of the discrimination power of the network and the volumetric features, and the comparison of the performance of different classifiers for the network features.

### Comparison of Features

The data set was randomly partitioned into 20 training and test groups, each with 150 samples for training and 75 samples for test.

Five methods were tested in the experiment: i) FN: the proposed method, using the four-layer hierarchical network features; ii) SN: using only the network features from the bottommost layer; iii) FN-NC: using the network features from all the four layers, but removing the edges across different layers; iv) SV: using the volumetric features from the bottommost layer; v) FV: using volumetric measures from all four layers.

The results are summarized in Table 6. The classification accuracy is averaged across the 20 randomly partitioned training and test groups. A paired *t*-test is conducted between the proposed method (FN) and the other four methods, respectively, to demonstrate the advantage of the proposed method. The *p*-value of the paired *t*-test is also reported. It can be seen from Table 6 that the proposed method (FN) is always statistically better (at the significance level of 0.05) than any of the other four methods.

**Table 1.6 Comparison of discrimination efficacy of features.**

-	Mean Test Accuracy (%)	Paired t-test p-value
FN	85.07	-
SN	83.00	0.00272
FN-NC	83.13	0.00367
SV	81.93	0.00166
FV	81.47	0.00015

From Table 1.6, the following results are observed:

- The proposed hierarchical network features in FN outperform the conventional volumetric features in SV. The advantage may come from using both regional interactions and the hierarchical structure.
- The better performance of SN over SV, and FN over FV demonstrate the benefits purely from using the regional interactions.
- The better performance of FN over SN demonstrates the benefit purely from the hierarchy. The advantage of the four-layer structure is statistically significant over the single-layer. Moreover, the result that FN statistically outperforms FN-NC indicates the necessity of using the cross-layer edges in the network.

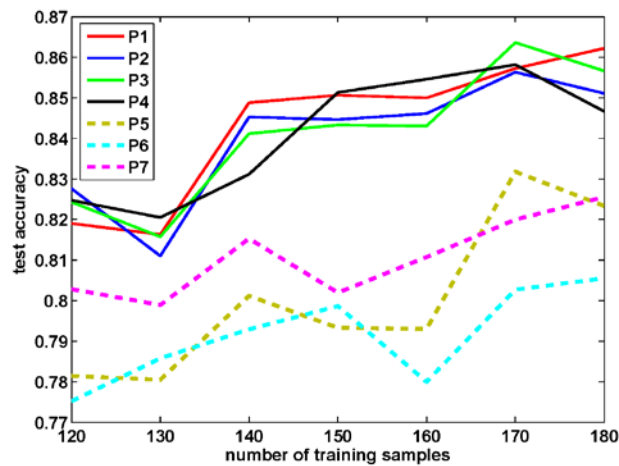
### Comparison of Classifiers

The classification performance of the proposed classification scheme was compared with other six possible schemes shown in Table 1.7. To simplify the description, the proposed scheme was denoted as P1, while the other six schemes in comparison were denoted as P2 ~ P7. To keep consistent with P1, each of the six schemes P2 ~ P7 was also divided into four steps: rough feature selection, refined feature selection, feature embedding and classification. Note that the first step, rough feature selection, was the same for all schemes P1 ~ P7.

The classification results are summarized in Fig.6 and Table 7. The overall classification accuracy in Table 7 is an average of accuracies at different numbers of training samples in Fig. 6. The best overall classification accuracy of 84.35% is obtained by the proposed scheme P1: VIP selection + PLS embedding + a linear SVM. This is slightly better than P2, where a nonlinear SVM is used. It can be seen that the classification schemes with PLS embedding (P1 ~ P4) achieve an overall accuracy above 84.0%, better than those without PLS embedding (P5 ~ P7). The supervised embedding methods, i.e., PLS (P1 ~ P4) and KFDA (P7), perform better than the unsupervised Laplacian Eigenmap embedding (P6). Moreover, PLS embedding (P1 ~ P4) preserves more discrimination than the nonlinear supervised embedding of KFDA (P7).

**Table 1.7 Configuration of classification schemes.**

Scheme	Configuration	Accuracy (%)
P1	VIP selection + PLS embedding + linear SVM	84.35
P2	VIP selection + PLS embedding + nonlinear SVM	84.03
P3	no selection + PLS embedding + linear SVM	84.11
P4	no selection + PLS embedding + nonlinear SVM	84.10
P5	SVM-RFE selection + no embedding + nonlinear SVM	80.07
P6	no selection + Laplacian Eigenmap embedding + nonlinear SVM	79.16
P7	no selection + KFDA embedding + linear SVM	81.08



**Fig. 1.6 Comparison of seven classification schemes on network features. The classification accuracy at each number of training samples is averaged over 20 randomly partitioned training and test groups. The scheme configurations are shown in Table 1.7.**

### Spatial Patterns

The most discriminative features selected by more than half of the training and test groups using the two-step feature selection are shown in Table 1.8. Note that each network feature characterizes the relationship between two ROIs, instead of an individual ROI. Table 1.8 is separated into two parts. On the upper portion of the table, the two ROIs of a network feature may be both associated with the MCI diagnosis, such as hippocampus, entorhinal cortex, fornix, cingulate etc, as reported in the literature (Cuingnet et al. 2011; Pengas et al. 2010). But their change speeds are different over two different groups. A typical example is the correlation between hippocampus and ventricle. It is known that the enlargement of ventricle is a biomarker for the diagnosis of the AD (Nestor et al. 2008). However, different from the hippocampus volume loss that often occurs at the very early stage of the dementia, the ventricle enlargement often appears in the middle and late stages.

On the lower portion of the table, the first ROI is associated with the disease, while the second ROI is constant to the disease. For example, it has been reported that the anterior and posterior limbs of internal capsule and the occipital lobe white matter are not significantly different between MCI and normal controls in a DTI study (Bozzali et al. 2002). As shown, the different progression patterns makes the correlation between the two regions the discriminative feature.

**Table 1.8 Selected discriminative features.**

hippocampus - amygdala
hippocampus - lingual gyrus
hippocampus - uncus
hippocampus - prefrontal/superolateral frontal lobe
hippocampus - globus palladus
hippocampus - entorhinal cortex
hippocampus - cingulate region
hippocampus - ventricle
hippocampus and amygdala and fornix - ventricle
uncus - fornix
hippocampus - posterior limb of internal capsule
globus palladus - anterior limb of internal capsule
hippocampus - occipital lobe WM

### 1.4.2 Multimodality-Based Diagnosis and Prognosis

We performed various experiments on the multimodality data obtained from the ADNI database. Only ADNI subjects with all corresponding MRI, PET, and CSF data at baseline were included. This yields a total of 202 subjects, including 51 AD patients, 99 MCI patients, and 52 normal controls (NCs). Table 1.9 lists the subject characteristics.

**Table 1.9 Demographic information of the subjects involved in the study.**

-	AD (n=51, 18F/33M)			MCI (n=99, 32F/67M)			NC (n=52, 18F/34M)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age	75.2	7.4	59~88	75.3	7.0	55~89	75.3	5.2	62~85
Education	14.7	3.6	4~20	15.9	2.9	8~20	15.8	3.2	8~20
MMSE	23.8	2.0	20~26	27.1	1.7	24~30	29.0	1.2	25~30
CDR	0.7	0.3	0.5~1	0.5	0.0	0.5~0.5	0.0	0.0	0~0

Image pre-processing was performed for all MRI and PET images. Specifically, we performed anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR

images into three different tissues: GM, WM, and CSF. With atlas warping, we partitioned each subject image into 93 ROIs. For each of the 93 ROIs, we computed the GM tissue volume from the subject's MRI image. For each PET image, we first rigidly aligned it with its respective MRI image, and then computed the average value of PET signals in each ROI. Therefore, for each subject, we obtained in total 93 features from its MRI image, 93 features from its PET image, and 3 features ( $A\beta_{42}$ ,  $t$ -tau and  $p$ -tau) from the CSF biomarkers.

#### 1.4.2.1 AD/MCI Diagnosis Using Multi-Kernel SVM

A 10-fold cross-validation strategy was used to compute the classification accuracy, as well as the sensitivity and the specificity. Specifically, the whole set of subject samples were equally partitioned into 10 subsets, and each time the subject samples within one subset were selected as the testing samples and all remaining subject samples in the 9 other subsets were used for training the multiple-kernel classifier. This process was repeated for 10 times to avoid any bias introduced by the random partitioning of the dataset during cross-validation. The SVM classifier was implemented using LIBSVM toolbox (Chang and Lin 2001), with a linear kernel and a default value for the parameter  $C$  (i.e.,  $C=1$ ). The weights in the multiple-kernel classification method were learned based on the training samples through a grid search in the range of 0 to 1 with a step size of 0.1. Specifically, we performed another 10-fold cross-validation on the training sample subset to determine the optimal values for the weights. For each feature  $f_i$  in the training samples, a common feature normalization scheme was adopted, i.e.,  $f_i = (f_i - \bar{f}_i) / \sigma_i$ , where  $\bar{f}_i$  and  $\sigma_i$  are respectively the mean and standard deviation of the  $i$ -th feature across all training samples. The estimated  $\bar{f}_i$  and  $\sigma_i$  will be used to normalize the corresponding feature of each test sample.

#### Multimodality Classification Based on MRI, PET, and CSF

We first tested the performance of our multimodality classification method in identification of AD (or MCI) from normal controls, based on MRI, PET, and CSF biomarkers of 202 baseline subjects in ADNI. Table 1.10 shows the classification rate of the multimodality classification method, compared with the methods using each individual modality only. Note that Table 1.10 shows only the averaged results of 10 independent experiments, along with the minimal and maximal values given in brackets. As can be observed from Table 1.10, the combined measurements of MRI, PET, and CSF consistently achieve more accurate discrimination between AD (or MCI) patients and normal controls. Specifically, for classifying AD from normal controls, the multimodality classification method can achieve a classification accuracy of 93.2%, a sensitivity of 93.0%, and a specificity of

93.3%, while the best accuracy on individual modality is only 86.5% (when using PET). On the other hand, for classifying MCI from normal controls, the multimodality classification method achieve a classification accuracy of 76.4%, a sensitivity of 81.8%, and a specificity of 66.0%, while the best accuracy on individual modality is only 72.0% (when using MRI).

For comparison with other multimodality classification methods, we also performed the use of direct feature concatenation as a baseline method for multimodality AD (or MCI) classification. Specifically, for each subject, we first concatenated 93 features from MRI, 93 features from PET, and 3 features from CSF, into a 189 dimensional vector. Remember that each feature has been normalized to have zero mean and unit standard deviation. Then, we performed SVM-based classification on all samples with a 10-fold cross-validation strategy as described above, and the classification results are shown in the bottom row of Table 1.10. As can be observed from Table 1.10, our kernel combination method consistently outperforms the baseline method for every performance measure.

**Table 1.10 Comparison of performance of single-modal and multimodal classification methods. The numbers in each bracket denote the minimal and maximal classification rate in 10 independent experiments.**

Method	AD vs. NC			MCI vs. NC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
MRI	86.2	86.0	86.3	72.0	78.5	59.6
	(82.9~89.0)	(82.7~88.7)	(83.1~89.1)	(68.4~74.7)	(75.6~80.6)	(55.1~63.7)
CSF	82.1	81.9	82.3	71.4	78.0	58.8
	(80.0~84.9)	(80.0~84.7)	(80.0~85.1)	(68.2~73.3)	(75.6~79.4)	(54.3~61.7)
PET	86.5	86.3	86.6	71.6	78.2	59.3
	(82.9~90.5)	(82.7~90.3)	(83.1~90.6)	(67.4~74.7)	(75.0~80.6)	(52.9~63.7)
Combined	93.2	93.0	93.3	76.4	81.8	66.0
	(89.0~96.5)	(88.7~96.3)	(89.1~96.6)	(73.5~79.7)	(79.4~84.4)	(62.6~70.3)
Baseline	91.5	91.4	91.6	74.5	80.4	63.3
	(88.5~96.5)	(88.3~96.3)	(88.6~96.6)	(71.9~78.2)	(78.3~83.3)	(59.7~68.3)

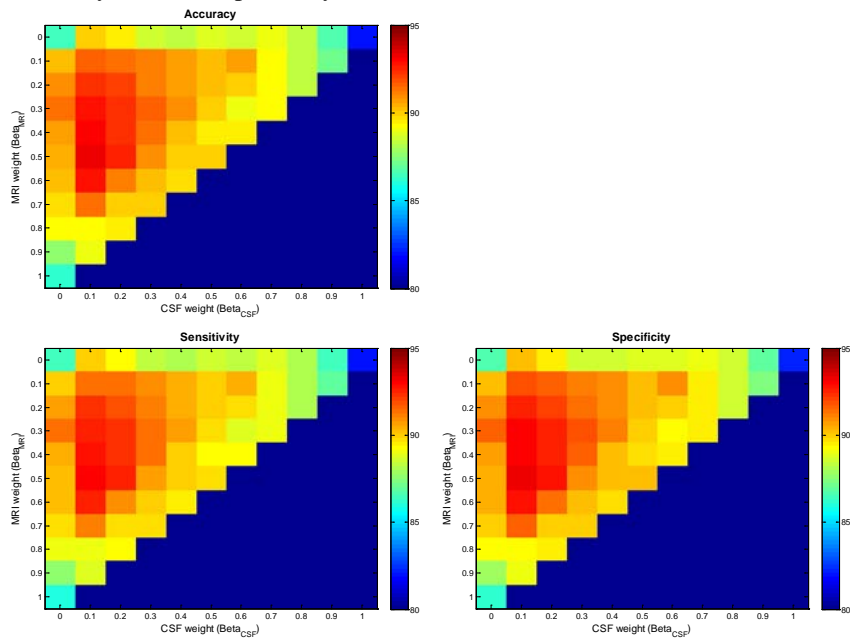
AD = Alzheimer's Disease, MCI = Mild Cognitive Impairment, NC = normal control, ACC = classification ACCuracy, SEN = SENSitivity, SPE = SPEcificity

### Comparison of Different Combination Schemes

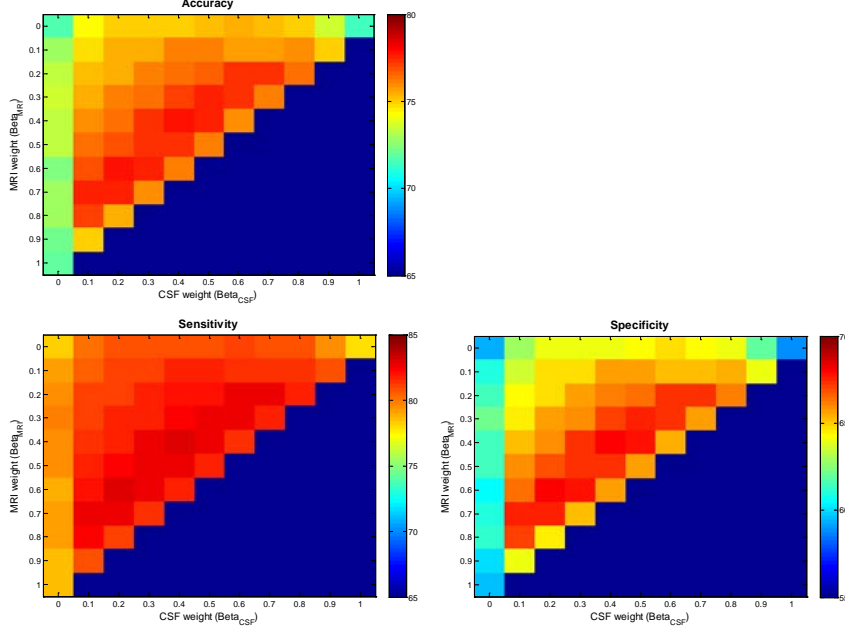
To investigate the effect of different combining weights, i.e.,  $\beta_{MRI}$ ,  $\beta_{CSF}$ ,  $\beta_{PET}$ , on the performance of our multimodality classification method, we tested all of their possible values, ranging from 0 to 1 at a step size of 0.1, under the constraint of ( $\beta_{MRI} + \beta_{CSF} + \beta_{PET} = 1$ ). Figs. 1.7 and 1.8 show the classification results, including accuracy (top row), sensitivity (bottom left), and specificity (bottom right), with respect to different combining weights of MRI, PET, and CSF. Note that, in each

subplot, only the squares in the upper triangular part have valid values because of the constraint ( $\beta_{\text{MRI}} + \beta_{\text{CSF}} + \beta_{\text{PET}} = 1$ ). For each plot, the three vertices of the upper triangle, i.e., the top left, top right, and bottom left squares, denote individual-modality based classification results using only PET ( $\beta_{\text{PET}} = 1$ ), CSF ( $\beta_{\text{CSF}} = 1$ ), and MRI ( $\beta_{\text{MRI}} = 1$ ), respectively.

As we can observe from Figs. 1.7 and 1.8, nearly all inner squares of the upper triangle have larger values (better classification) than the three vertices, which demonstrates the effectiveness of combining three modalities in AD (or MCI) classification. Moreover, for most plots, there are a substantially large set of squares with higher classification accuracy. Further observation indicates that the squares with higher accuracy appear mainly in the inner of each triangle, instead of the boundary, implying that each modality is indispensable for achieving good classification. Similar to what we have observed from Table 1.10, Figs. 7 and 8 also show that, for AD classification, the differences among accuracy, sensitivity, and specificity are small, while, for MCI classification, it tends to have a higher sensitivity but lower specificity.



**Fig. 1.7** AD Classification results with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint: ( $\beta_{\text{MRI}} + \beta_{\text{CSF}} + \beta_{\text{PET}} = 1$ ). Note that for each plot, the top left, top right, and bottom left squares denote the individual-modality based classification results using PET ( $\beta_{\text{PET}} = 1$ ), CSF ( $\beta_{\text{CSF}} = 1$ ), and MRI ( $\beta_{\text{MRI}} = 1$ ), respectively.



**Fig. 1.8** MCI Classification results with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint:  $(\beta_{\text{MRI}} + \beta_{\text{CSF}} + \beta_{\text{PET}} = 1)$ . Note that for each plot, the top left, top right, and bottom left squares denote the individual-modality based classification results using PET ( $\beta_{\text{PET}}=1$ ), CSF ( $\beta_{\text{CSF}}=1$ ), and MRI ( $\beta_{\text{MRI}}=1$ ), respectively..

#### 1.4.2.2 Diagnosis Using Semi-Supervised Multimodality Classification

We compared mLapRLS with mRLS on the multimodality (MRI, PET, and CSF) data. Specifically, a 10-fold cross-validation was performed on 51 AD patients and 52 NC subjects to get the labeled training data and testing data. Unlabeled data were obtained from those 99 MCI subjects. A linear kernel was used for both algorithms. Following (Belkin et al. 2006), for mRLS, we set the parameters as  $\gamma_A=0.05/l$  and  $\gamma_B=0$ ; for mLapRLS, we set  $\gamma_A=0.05/l$  and  $\gamma_B=0.05(l+u)^2/l$ . Here,  $l$  denotes the number of AD and NC subjects, and  $u$  is the number of MCI subjects. The Euclidean distance is used for each modality in Eqn. (1.8). For both algorithms, the values of the weighting parameters  $\beta_{ms}$  are determined through cross-validation using grid search.

Fig. 1.9 shows the classification results of both algorithms on the multimodality data, which include classification accuracy, sensitivity, specificity, and AUC. The results in Fig. 1.9 indicate that, by using the MCI subjects as additional unlabeled data, mLapRLS significantly improves the performances of distinguishing AD

from NC subjects, compared to those by mRLS that uses only AD and NC subjects as samples for training classifier. Specifically, the AUC values of mLapRLS and mRLS are 0.985 and 0.946, respectively. The results validate the effectiveness of mLapRLS in using additional data (i.e., MCI subjects) to enhance the AD classification.

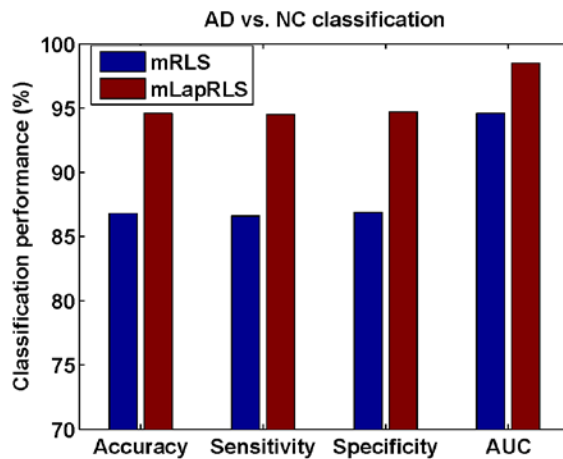


Fig. 1.9 Classification results on multimodality data.

Finally, in Fig. 1.10, we show the classification accuracy of the mLapRLS algorithm with respect to different number of MCI subjects used for helping training. As we can see from Fig. 1.10, as the number of included MCI subjects increases, the classification accuracy of mLapRLS also steadily increases, which again validates the usefulness of using MCI subjects for helping classification between AD and NC.

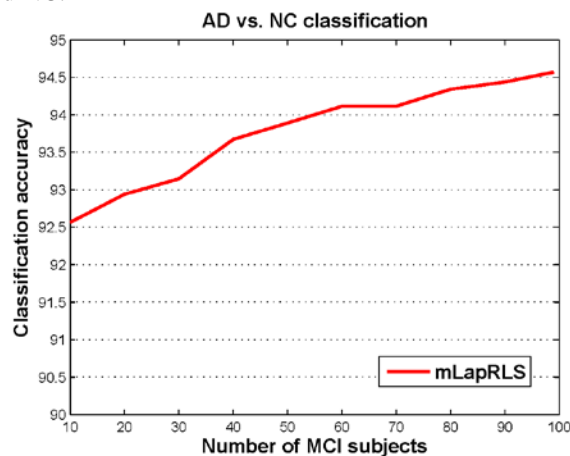


Fig. 1.10 Classification accuracy with respect to the different number of MCI subjects used to help train the multimodality classifier.

## 1.5 Summary

In the past two decades, machine learning techniques have been proven to be important for effective neurodegenerative disorders diagnosis and prognosis, particularly for AD and MCI. Essentially, machine learning techniques that have been applied for AD and MCI diagnosis and prognosis can be categorized into single modality and multimodality based approaches. Some recent developments in this area have been discussed in this chapter. In single modality based approaches, information on local microstructural characteristics of water diffusion (DTI), hemodynamic response related to neural activity (fMRI) and structural atrophy (T1-weighted imaging) is extracted using connectivity networks to provide a comprehensive representation of brain alterations for improved classification performance. For DTI, a collection of physiological parameters are derived along the tracked fibers for better characterization of brain circuitry. The multi-spectral characterization provides a localized analysis of BOLD signals by decomposing the frequency interval into several sub-bands. For T1-weighted images, hierarchical brain connectivity networks derived from the structural images provides a more effective way of characterizing subtle changes than by using local isolated measures. It is widely accepted that different modalities can convey complementary information and is useful for AD and MCI diagnosis and prognosis. Based on this observation, many machine learning techniques have been applied to integrate information from multiple modalities. Multi-kernel SVM, when used to integrate complementary information from structural MRI, PET and CSF, demonstrates significant improvements in AD and MCI diagnosis and prognosis. As a remedy to small sample size problem, a semi-supervised learning technique is introduced to derive additional information from MCI data for improving the discriminative power of the constructed classifiers. The increased accuracy, sensitivity, and specificity of these approaches indicate that machine learning techniques are a viable alternative to clinical diagnosis of brain alterations associated with cognitive impairment.

## References

- Achard S, Bassett DS, Meyer-Lindenberg A, Bullmore E (2008) Fractal connectivity of long-memory networks. *Physical Review E* 77:036104
- Bach FR, Lanckriet GRG, Jordan MI (2004) Multiple kernel learning, conic duality, and the SMO algorithm. Paper presented at the Proceedings of the twenty-first international conference on Machine learning (ICML'04),
- Bankman IN (2008) *Handbook of Medical Image Processing and Analysis*. Academic Press,
- Bassett DS, Bullmore E (2006) Small-World Brain Networks. *The Neuroscientist* 12 (6):512 - 523

- Belkin M, Niyogi P, Sindhvani V, Bartlett P (2006) Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research* 7:2399 - 2434
- Bischkopf J, Busse A, Angermeyer MC (2002) Mild Cognitive Impairment - A Review of Prevalence, Incidence and Outcome According to Current Approaches. *Acta Psychiatrica Scandinavica* 106:403 - 414
- Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine* 34 (4):537 - 541
- Bozzali M, Falini A, Franceschi M, Cercignani M, Zuffi M, Scotti G, Comi G, Filippi M (2002) White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. *J Neurol Neurosurg Psychiatry* 72 (6):742 - 746
- Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's dDisease. *Alzheimer's & Dementia* 3 (3):186 - 191
- Chang C-C, Lin C-J (2001) LIBSVM: a library for support vector machines.
- Chapelle O, Scholkopf B, Zien A (2006) *Semi-Supervised Learning*. MIT Press, Cambridge, MA
- Chetelat G, Baron JC (2003) Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage* 18 (2):525 - 541
- Chetelat G, Eustache F, Viader F, De La Sayette V, Pelerin A, Mezenge F, Hannequin D, Dupuy B, Baron JC, Desgranges B (2005) FDG-PET measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment. *Neurocase* 11 (1):14 - 25
- Chu C, Kloppel SS, Draganski CSB, Jack Jr. C, Ashburner J, Frackowiak RSJ (2007) Regression analysis for clinical scores of Alzheimer's Disease using multivariate machine learning method. Paper presented at the Human Brain Mapping 2007, Chicago,
- Chupin M, Gerardin E, Cuingnet R, Boutet C, Lemieux L, Lehericy S, Benali H, Garnero L, Colliot O, Alzheimer's Disease Neuroimaging Initiative (2009) Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19 (6):579 - 587
- Colliot O, Chetelat G, Chupin M, Desgranges B, Magnin B, Benali H, Dubois B, Garnero L, Eustache F, Lehericy S (2008) Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology* 248 (1):194 - 201
- Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME (2001) Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am J Neuroradiology* 22:1326 - 1333

- Courchesne E, Chisum HJ, Townsend J, Cowles A, Covington J, Egaas B, Harwood M, Hinds S, Press GA (2000) Normal brain development and aging: Quantitative analysis at in Vivo MR imaging in healthy volunteers. *Radiology* 216:672 - 682
- Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert M-O, Chupin M, Benali H, Colliot O, The Alzheimer's Disease Neuroimaging Initiative (2011) Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage* 56 (2):766 - 781
- Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2010) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging*
- Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobio Aging* 29:514 - 523
- Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, Grodstein F, Wright CI, Blacker D, Rosas HD, Sperling RA, Atri A, Growdon JH, Hyman BT, Morris JC, Fischl B, Buckner RL (2009) The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb Cortex* 19 (3):497 - 510
- Diehl J, Grimmer T, Drzezga A, Riemenschneider M, Forstla H, Kurz A (2004) Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. *Neurobio Aging* 25 (8):1051 - 1056
- Duchesne S, Caroli A, Geroldi C, Frisoni GB, Collins DL (2005) Predicting clinical variable from MRI features: application to MMSE in MCI. *Med Image Comput Assist Interv* 8 (1):392 - 399
- Fan Y, Batmanghelich N, Clark CM, Davatzikos C, the Alzheimer's Disease Neuroimaging Initiative (2008a) Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 39:1731 - 1743
- Fan Y, Gur RE, Gur RC, Wu X, Shen D, Calkins ME, Davatzikos C (2008b) Unaffected family members and Schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study. *Biol Psychiatry* 63 (1):118 - 124
- Fan Y, Kaufner D, Shen D (2009) Estimating clinical variables from brain images using Bayesian regression. *International Conference on Alzheimer's Disease*
- Fan Y, Resnick SM, Wu X, Davatzikos C (2008c) Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. *Neuroimage* 41:277 - 285

- Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C (2007) COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements. *IEEE T Medical Imaging* 26 (1):93 - 105
- Fellgiebel A, Scheurich A, Bartenstein P, Muller MJ (2007) FDG-PET and CSF phospho-tau for prediction of cognitive decline in mild cognitive impairment. *Psychiatry Research: Neuroimaging* 155 (2):167 - 171
- Fitzpatrick JM, Sonka M (2000) Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis, vol PM80SC. SPIE-The International Society for Optical Engineering,
- Fjell AM, Walhovd KB, Fennema-Notestine C, McEvoy LK, Hagler DJ, Holland D, Brewer JB, Dale AM, the Alzheimer's Disease Neuroimaging Initiative (2010) CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's Disease. *The Journal of Neuroscience* 30 (6):2088 - 2101
- Friston KJ, Frith CD, Liddle PF, Frackowiak RS (1993) Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow & Metabolism* 13:5 - 14
- Geroldi C, Rossi R, Calvagna C, Testa C, Bresciani L, Binetti G, Zanetti aO, Frisoni GB (2006) Medial temporal atrophy but not memory deficit predicts progression to dementia in patients with mild cognitive impairment. *Journal Neurol Neurosurg Psychiatry* 77:1219 - 1222
- Gong G, He Y, Concha L, Lebel C, Gross DW, Evans AC, Beaulieu C (2009) Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb Cortex* 19:524 - 536
- Greicius M (2008) Resting-state functional connectivity in neuropsychiatric disorders. *Current Opinion in Neurology* 21:424 - 430
- Greicius MD, Srivastava G, Reiss AL, Menon V (2004) Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *PNAS* 101 (13):4637 - 4642
- Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, al. e (2004) Mild cognitive impairment can be distinguished from Alzheimer's disease and normal aging for clinical trials. *Arch Neurol* 61 (1):59 - 66
- Guyon I, Weston J, Barnhill S, Vapnik V (2004) Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46 (1 - 3):389 - 422
- Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008) Mapping the structural core of human cerebral cortex. *PLoS Computational Biology* 6:e159
- Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC, the Alzheimer's Disease Neuroimaging Initiative (2009a) Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48 (1):138 - 149
- Hinrichs C, Singh V, Xu G, Johnson S (2009b) MKL for Robust Multi-modality AD Classification. *Med Image Comput Comput Assist Interv* 2009: Part II:786 - 794

- Jack Jr. CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* 9 (1):119 - 128
- Jack Jr. CR, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, Tangalos EG, E. Kokmen E (1998) Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* 51 (4):993 - 999
- Jack Jr. CR, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC (2005) Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI. *Neurology* 65 (8):1227 - 1231
- Jiang J, Trundle P, Ren J (2010) Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics* 34 (8):617 - 631
- Johnson SC, Schmitz TW, Moritz CH, Meyerand ME, Rowley HA, Alexander AL, Hansen KW, Gleason CE, Carlsson CM, Ries ML, Asthana S, Chen K, Reiman EM, Alexander GE (2006) Activation of brain regions vulnerable to Alzheimer's disease: The effect of mild cognitive impairment. *Neurobio Aging* 27 (11):1604 - 1612
- Karas GB, Burton EJ, Rombouts SARB, van Schijndel RA, O'Brien JT, McKeith IG, Williams D, Ballard C, Barkhof F (2003) A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. *Neuroimage* 18 (4):895 - 907
- Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack Jr CR, Ashburner J, Frackowiak RSJ (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain* 131 (3):681 - 689
- Lanckriet GR, Deng M, Cristianini N, Jordan MI, Noble WS (2004a) Kernel-based data fusion and its application to protein function prediction in yeast. Paper presented at the Pac Symp Biocomput,
- Lanckriet GR, Cristianini N, Bartlett P, Ghaoui LE, Jordan MI (2004b) Learning the Kernel Matrix with Semidefinite Programming. *Journal of Machine Learning Research* 5:27 - 72
- Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack Jr. CR, Weiner MW, Jagust WJ, Alzheimer's Disease Neuroimaging Initiative (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 75 (3):230 - 238
- Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C (2004) Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage* 21:46 - 57
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34 (7):939 - 944

- Morra JH, Tu Z, Apostolova LG, Green AE, Toga AW, Thompson PM (2010) Comparison of AdaBoost and support vector machines for detecting Alzheimer's Disease through automated hippocampal segmentation. *IEEE T Med Imaging* 29 (1):30 - 43
- Nestor PJ, Scheltens P, Hodges JR (2004) Advances in the early detection of Alzheimer's disease. *Nature* 5
- Nestor SM, Rupsingh R, Borrie M, Smith M, Accomazzi V, Wells JL, Fogarty J, Bartha R, Alzheimer's Disease Neuroimaging Initiative t (2008) Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's Disease Neuroimaging Initiative database. *Brain* 131 (9):2443 - 2454
- Pengas G, Hodges JR, Watson P, Nestor PJ (2010) Focal posterior cingulate atrophy in incipient Alzheimer's disease. *Neurobio Aging* 31 (1):25 - 33
- Rakotomamonjy A (2003) Variable selection using SVM based criteria. *Journal of Machine Learning Research: Special issue on special feature* 3:1357 - 1370
- Rose SE, Janke AL, Chalk JB (2007) Gray and white matter changes in Alzheimer's disease: A diffusion tensor imaging study. *Journal of Magnetic Resonance Imaging* 27 (1):20 - 26
- Rosipal R, Kramer N (2006) Overview and recent advances in partial least squares. *Lecture Notes in Computer Science* 3940:34 - 51
- Rubinov M, Sporns O (2010) Complex networks measures of brain connectivity: Uses and interpretations. *Neuroimage* 52 (3):1059 - 1069
- Scholkopf B, Smola AJ (2002) *Learning with Kernels*. The MIT Press,
- Sorg C, Riedl V, Muhlau M, Calhoun VD, Eichele T, Laer L, Drzezga A, Forstl H, Kurz A, Zimmer CZ, M. WA (2007) Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *PNAS* 104 (47):18760 - 18765
- Sporns O, Zwi JD (2004) The small world of the cerebral cortex. *Neuroinformatics* 2:145 - 161
- Thompson PM, Apostolova LG (2007) Computational anatomical methods as applied to ageing and dementia. *British Journal of Radiology* 80:S78 - S91
- Thompson PM, Hayashi KM, de Zubicaray G, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doddrell DM, Toga AW (2003) Dynamics of gray matter loss in Alzheimer's Disease. *The Journal of Neuroscience* 23 (3):994 - 1005
- Thompson PM, Hayashi KM, Sowell ER, Gogtay N, Giedd JN, Rapoport JL, de Zubicaray GI, Janke AL, Rose SE, Semple J, Doddrell DM, Wang Y, van Erp TG, Cannon TD, Toga AW (2004) Mapping cortical change in Alzheimer's disease, brain development, and Schizophrenia. *Journal of Neuroscience* 23:S2 - S18
- Thompson PM, Mega MS, Woods RP, Zoumalan CI, Lindshield CJ, Blanton RE, Moussai J, Holmes CJ, Cummings JL, Toga AW (2001) Cortical changes in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb Cortex* 11 (1):1 - 16

- Tiwari P, Kurhanewicz J, Rosen M, Madabhushi A (2010) Semi Supervised Multi Kernel (SeSMiK) Graph Embedding: Identifying Aggressive Prostate Cancer via Magnetic Resonance Imaging and Spectroscopy. *Med Image Comput Assist Interv* 2010:666 - 673
- Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010) Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties and Optimization. *J Neurophysiol* 103:297 - 321
- Vapnik VN (1999) *The nature of statistical learning theory* (Statistics for Engineering and Information Science). Springer-Verlag,
- Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack Jr CR (2008) Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *Neuroimage* 39 (3):1186 - 1197
- Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack J, C. R., On behalf of the Alzheimer's Disease Neuroimaging Initiative (2009) MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology* 73 (4):294 - 301
- Walhovd KB, Fjell aAM, Dale AM, McEvoy LK, Brewer J, Karow DS, Salmon DP, Fennema-Notestine C, Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler Jr. DJ, Jennings RG, Karow D, Dale AM, Alzheimer's Disease Neuroimaging Initiative (2010a) Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *Am J Neuroradiol* 31 (2):347 - 354
- Walhovd KB, Fjell AM, Dale AM, McEvoy LKM, Brewer J, Karow DS, Salmon DPS, Fennema-Notestine C, Alzheimer's Disease Neuroimaging Initiative (2010b) Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobio Aging* 31 (7):1107 - 1121
- Wang Z, Chen S, Sun T (2008) MultiK-MHKS: A Novel Multiple Kernel Learning Algorithm. *IEEE T Pattern Analysis and Machine Intelligence* 30 (2):348 - 353
- Watts DJ, Strogatz SH (1998) Collective dynamics of `small-world' networks. *Nature* 393:440 - 442
- Wee C-Y, Yap P-T, Li W, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D (2011a) Enriched White Matter Connectivity Networks for Accurate Identification of MCI Patients. *Neuroimage* 54 (3):1812 - 1822
- Wee C-Y, Yap P-T, Wang L, Shen D Classification of MCI patients via functional connectivity networks. In: *ISMRM' 2011, Quebec, Canada, May 2011b*.
- Westman E, Simmons A, Zhang Y, Muehlboeck J-S, Tunnard C, Liu Y, Collins L, Evans A, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Wahlund L-O, for the AddNeuroMed Consor-

- tium (2010) Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage* 54 (2):1178 - 1187
- Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, Jack Jr. CR (2007) 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain* 130 (7):1777 - 1786
- Whitwell JL, Shiung MM, Przybelski S, Weigand SD, Knopman DS, Boeve BF, Peterson RC, Jack Jr. CR (2008) MRI patterns of atrophy associated with progression to AD in amnesic Mild Cognitive Impairment. *Neurology* 70 (7):512 - 520
- Wold S, Eriksson L, Trygg J, Kettaneh N (1993) PLS - Partial Least- Squares Projections to Latent Structures. *3D QSAR in Drug Design: Volume 1: Theory Methods and Applications* 1:523 - 550
- Ye J, Chen K, Wu T, Li J, Zhao Z, Patel R, Bae M, Janardan RI, Liu H, Alex G, Reiman EI (2008) Heterogeneous data fusion for Alzheimer's disease study. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*
- Zhang D, Shen D (2011) Semi-supervised multimodal classification of Alzheimer's disease. *IEEE International Symposium on Biomedical Imaging (ISBI'11)*
- Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Alzheimer's Disease Neuroimaging Initiative t (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55 (3):856 - 867
- Zhang Y, Schuff N, Jahng G-H, Bayne W, Mori S, Schad L, Mueller S, Du A-T, Kramer JH, Yaffe K, Chui H, Jagust WJ, Miller BL, Weiner MW (2007) Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. *Neurology* 68 (1):13 - 19
- Zhou L, Wang Y, Li Y, Yap P-T, Shen D Hierarchical anatomical brainnetworks for MCI prediction by partial least square analysis. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2011.*