

Published in final edited form as:

Neuroimage. 2011 April 1; 55(3): 856–867. doi:10.1016/j.neuroimage.2011.01.008.

Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment

Daoqiang Zhang^a, Yaping Wang^{a,b}, Luping Zhou^a, Hong Yuan^a, Dinggang Shen^a, and Alzheimer's Disease Neuroimaging Initiative¹

^aDepartment of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599

^bDepartment of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi Province, China

Abstract

Effective and accurate diagnosis of Alzheimer's disease (AD), as well as its prodromal stage (i.e., mild cognitive impairment (MCI)), has attracted more and more attentions recently. So far, multiple biomarkers have been shown sensitive to the diagnosis of AD and MCI, i.e., structural MR imaging (MRI) for brain atrophy measurement, functional imaging (e.g., FDG-PET) for hypometabolism quantification, and cerebrospinal fluid (CSF) for quantification of specific proteins. However, most existing research focuses on only a single modality of biomarkers for diagnosis of AD and MCI, although recent studies have shown that different biomarkers may provide complementary information for diagnosis of AD and MCI. In this paper, we propose to combine three modalities of biomarkers, i.e., MRI, FDG-PET, and CSF biomarkers, to discriminate between AD (or MCI) and healthy controls, using a kernel combination method. Specifically, ADNI baseline MRI, FDG-PET, and CSF data from 51 AD patients, 99 MCI patients (including 43 MCI converters who had converted to AD within 18 months and 56 MCI non-converters who had not converted to AD within 18 months), and 52 healthy controls are used for development and validation of our proposed multimodal classification method. In particular, for each MR or FDG-PET image, 93 volumetric features are extracted from the 93 regions of interest (ROIs), automatically labeled by an atlas warping algorithm. For CSF biomarkers, their original values are directly used as features. Then, a linear support vector machine (SVM) is adopted to evaluate the classification accuracy, using a 10-fold cross-validation. As a result, for classifying AD from healthy controls, we achieve a classification accuracy of 93.2% (with a sensitivity of 93% and a specificity of 93.3%) when combining all three modalities of biomarkers, and only 86.5% when using even the best individual modality of biomarkers. Similarly, for classifying MCI from healthy controls, we achieve a classification accuracy of 76.4% (with a sensitivity of 81.8% and a specificity of 66%) for our combined method, and only 72% even using the best individual modality of biomarkers. Further analysis on MCI sensitivity of our combined method indicates that 91.5% of MCI converters and 73.4% of MCI non-converters are correctly classified. Moreover, we also evaluate the classification performance when employing a feature selection method to select the most discriminative MR and FDG-PET features. Again, our combined

© 2010 Elsevier Inc. All rights reserved.

¹Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

method shows considerably better performance, compared to the case of using an individual modality of biomarkers.

Keywords

Alzheimer's disease (AD); MCI; Multimodal classification; AD biomarkers; MRI; PET; CSF

Introduction

Alzheimer's disease (AD) is the most common form of dementia in elderly people worldwide. 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 (Ron et al., 2007). Thus, accurate diagnosis of AD, especially for its early stage also known as amnesic mild cognitive impairment (MCI), is very important. It is known that AD is related to the structural atrophy, pathological amyloid depositions, and metabolic alterations in the brain (Jack et al., 2010; Nestor et al., 2004). At present, several modalities of biomarkers have been proved to be sensitive to AD and MCI, including the brain atrophy measured in magnetic resonance (MR) imaging (de Leon et al., 2007; Du et al., 2007; Fjell et al., 2010; McEvoy et al., 2009), hypometabolism measured by functional imaging (De Santi et al., 2001; Morris et al., 2001), and quantification of specific proteins measured through CSF (Bouwman et al., 2007b; Fjell et al., 2010; Mattsson et al., 2009; Shaw et al., 2009).

However, most existing pattern classification methods just use one individual modality of biomarkers for diagnosis of AD or MCI, which may affect the overall classification performance. For example, many high-dimensional classification methods use only the structural MRI brain images for classification between AD (or MCI) and healthy controls (Cuingnet et al., 2010; Fan et al., 2008a; Fan et al., 2007; Gerardin et al., 2009; Klöppel et al., 2008; Lao et al., 2004; Magnin et al., 2009; Misra et al., 2009; Oliveira et al., 2010; Westman et al., 2010). Also, according to the features being extracted from the structural MRI, the existing classification methods can be roughly divided into three categories, using 1) voxel-wise tissue probability (Fan et al., 2007; Klöppel et al., 2008; Lao et al., 2004; Magnin et al., 2009), 2) cortical thickness (Desikan et al., 2009; Lerch et al., 2008; Oliveira et al., 2010; Querbes et al., 2009), and 3) hippocampal volumes (Gerardin et al., 2009; West et al., 2004). It was found that most effective features for AD or MCI classification are actually extracted from the atrophic regions, i.e., hippocampus, entorhinal cortex, parahippocampal gyrus, and cingulate, which are consistent with previous findings using group comparison methods (Chetelat et al., 2002; Convit et al., 2000; Fox and Schott, 2004; Jack et al., 1999; Misra et al., 2009). In addition to structural MRI, another important modality of biomarkers for AD or MCI detection is fluorodeoxyglucose positron emission tomography (FDG-PET) (Chetelat et al., 2003; Foster et al., 2007; Higdon et al., 2004). With FDG-PET, some recent studies have reported the reduction of glucose metabolism in parietal, posterior cingulate, and temporal brain regions for AD patients (Diehl et al., 2004; Drzezga et al., 2003). Besides these neuroimaging techniques, there are also some biological or genetic biomarkers developed for diagnosis of AD or MCI. For example, researchers have found 1) the increased CSF total tau (t-tau) and tau hyperphosphorylated at threonine 181 (p-tau) are related to the neurofibrillary tangle pathology, 2) the decreased amyloid β ($A\beta_{42}$) indicates amyloid plaque pathology, and 3) the presence of the apolipoprotein E (APOE) $\epsilon 4$ allele can predict cognitive decline or conversion to AD (Bouwman et al., 2007b; de Leon et al., 2007; Fjell et al., 2010; Ji et al., 2001).

Actually, different biomarkers provide complementary information, which may be useful for diagnosis of AD or MCI when used together (Apostolova et al., 2010; de Leon et al., 2007;

Fjell et al., 2010; Foster et al., 2007; Landau et al., 2010; Walhovd et al., 2010b). It was reported that FDG-PET and MRI measures are differentially sensitive to memory in health and disease (Walhovd et al., 2010b). A recent study also shows that the morphometric changes in AD and MCI are related to CSF biomarkers, but can also provide complementary information to CSF biomarkers (Fjell et al., 2010). A more recent study has compared the respective prognostic ability of genetic, CSF, neuroimaging, and cognitive measures obtained in the same participants, indicating that there exists complementary information among these biomarkers which may aid in the future diagnosis of AD and MCI (Landau et al., 2010). Inspired by these findings, a few studies have used two or more biomarkers simultaneously for detection of AD and MCI, i.e., using MRI and CSF in (Bouwman et al., 2007a; Vemuri et al., 2009), MRI and cognitive testing in (Geroldi et al., 2006; Visser et al., 2002), 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).

Although the use of multiple biomarkers yields promising results, the above methods may be limited. First, only a few manually selected brain regions are generally considered for MRI and PET based classification of AD or MCI. However, the structural and functional features measured from a limited set of pre-defined regions may be not able to reflect the spatial-temporal pattern of structural and physiological abnormalities in their entirety (Fan et al., 2008b). Second, most above methods are primarily designed to characterize group differences, not for individual classification. Although there exist some methods combining two modalities of biomarkers for individual classification, i.e., using both MRI and PET (Fan et al., 2008b; Hinrichs et al., 2009a; Hinrichs et al., 2009b; Ye et al., 2008), both MRI and CSF (Davatzikos et al., 2010), or both MRI and APOE biomarkers (Ye et al., 2008), there is still few method that combines all three modalities of biomarkers (MRI, PET, and CSF) for classification, which we will show the benefit of combining all three biomarkers for AD or MCI diagnosis in this paper.

Specifically, we will combine the measurements from all three biomarkers, i.e., MRI, PET, and CSF, to discriminate between AD and healthy controls, or between MCI and healthy controls. To effectively combine three different biomarkers for classification, we use a simple-while-effective multiple-kernel combination method. This method can be naturally embedded into the conventional SVM classifier without extra steps. Our experimental results show that the combination of different measurements from MRI, PET, and CSF demonstrates much better performance in AD or MCI classification, compared to the case of using even the best individual modality of biomarkers.

Methods

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early 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.

ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites

across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research – approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years (see www.adni-info.org for up-to-date information). The research protocol was approved by each local institutional review board and written informed consent is obtained from each participant.

Subjects

The ADNI general eligibility criteria are described at www.adni-info.org. Briefly, subjects are between 55-90 years of age, having a study partner able to provide an independent evaluation of functioning. Specific psychoactive medications will be excluded. General inclusion/exclusion criteria are as follows: 1) healthy subjects: Mini-Mental State Examination (MMSE) scores between 24-30, a Clinical Dementia Rating (CDR) of 0, non-depressed, non MCI, and nondemented; 2) MCI subjects: MMSE scores between 24-30, a memory complaint, having objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia; and 3) Mild AD: MMSE scores between 20-26, CDR of 0.5 or 1.0, and meets the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD.

In this paper, only ADNI subjects with all corresponding MRI, CSF and PET baseline data are included. This yields a total of 202 subjects including 51 AD patients, 99 MCI patients (43 MCI converters who had converted to AD within 18 months and 56 MCI non-converters who had not converted to AD within 18 months), and 52 healthy controls. Table 1 lists the demographics of all these subjects. Subject IDs are given in Supplemental Table 5.

MRI

All structural MR scans used in this paper were acquired from 1.5T scanners. Data were collected across a variety of scanners with protocols individualized for each scanner, as defined at www.loni.ucla.edu/ADNI/Research/Cores/index.shtml. Briefly, raw Digital Imaging and Communications in Medicine (DICOM) MRI scans were downloaded from the public ADNI site (www.loni.ucla.edu/ADNI), reviewed for quality, and automatically corrected for spatial distortion caused by gradient nonlinearity and B₁ field inhomogeneity.

PET

We downloaded the baseline PET data from the ADNI web site (www.loni.ucla.edu/ADNI) in December 2009. A detailed description of PET protocols and acquisition can be found at www.adni-info.org. Briefly, PET images were acquired 30-60 minutes post-injection, averaged, spatially aligned, interpolated to a standard voxel size, intensity normalized, and smoothed to a common resolution of 8-mm full width at half maximum.

CSF

We downloaded the baseline CSF A β ₄₂, t-tau and p-tau data from the ADNI web site (www.loni.ucla.edu/ADNI) in December 2009. The CSF collection and transportation protocols are provided in the ADNI procedural manual on www.adni-info.org. Briefly, CSF was collected in the morning after an overnight fast using a 20- or 24-gauge spinal needle, frozen within 1 hour of collection, and transported on dry ice to the ADNI Biomarker Core laboratory at the University of Pennsylvania Medical Center. In this study, CSF A β ₄₂, CSF t-tau and CSF p-tau are used as the features.

Image analysis

Image pre-processing is performed for all MR and PET images. First, we do anterior commissure (AC) – posterior commissure (PC) correction on all images, and use the N3 algorithm (Sled et al., 1998) to correct the intensity inhomogeneity. Next, we do skull-stripping on structural MR images using both brain surface extractor (BSE) (Shattuck et al., 2001) and brain extraction tool (BET) (Smith, 2002), followed by manual edition and intensity inhomogeneity correction. After removal of cerebellum, FAST in the FSL package (Zhang et al., 2001) is used to segment structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). After registration using HAMMER (Shen and Davatzikos, 2002), we obtain the subject-labeled image based on a template with 93 manually labeled ROIs (Kabani et al., 1998). For each of the 93 ROI regions in the labeled MR image, we compute the volume of GM tissue in that ROI region as a feature. For PET image, we first align it to its respective MR image of the same subject using a rigid transformation, and then compute the average intensity of each ROI region in the PET image as a feature. Therefore, for each subject, we totally obtain 93 features from MRI image, other 93 features from PET image, and 3 features from CSF biomarkers.

Multimodal data fusion and classification

A general framework based on kernel methods (Scholkopf and Smola, 2002) is presented here to combine multiple biomarkers (MRI, PET, and CSF) for discriminating between AD (or MCI) and healthy controls. This kernel-based method can be easily embedded into the conventional SVM classifier for high-dimensional pattern classification, without extra steps. Moreover, unlike other combining methods which can only process one type of data, i.e., numeric data type, our method can combine multiple types of data such as numeric data, string, and graph.

Before introducing the kernel combination method, we first briefly review the standard single-kernel SVM algorithm. The main idea of SVM is summarized as follows. First, the linearly nonseparable samples are mapped from their original space to a higher or even infinite dimensional feature space, where they are more likely to be linearly separable than in the original lower-dimensional space, through a kernel-induced implicit mapping function. Then, a maximum margin hyperplane is sought in the higher-dimensional space.

Now we will present the multiple-kernel SVM which can be used to integrate multiple modalities of biomarkers (i.e., MRI, PET and CSF) for individual classification of AD (or MCI) from healthy controls. 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:

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

Where $\mathbf{w}^{(m)}$, $\phi^{(m)}$ and $\beta_m \geq 0$ denote the normal vector of hyperplane, the kernel-induced mapping function, and the combining weight on the m -th modality, respectively.

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

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

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. The symbol n is the number of training samples.

For a new test sample $\mathbf{x} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(M)}\}$, we first denote

$k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) = \phi^{(m)}(\mathbf{x}_i^{(m)})^T \phi^{(m)}(\mathbf{x}_j^{(m)})$ 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:

$$f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(M)}) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_i \sum_{m=1}^M \beta_m k^{(m)}(\mathbf{x}_i^{(m)}, \mathbf{x}^{(m)}) + b \right).$$

It's easy to know that the 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 multimodal training samples \mathbf{x}_i and \mathbf{x}_j , and

$k(\mathbf{x}_j, \mathbf{x}) = \sum_m \beta_m k^{(m)}(\mathbf{x}_j^{(m)}, \mathbf{x}^{(m)})$ as a mixed kernel between the multimodal training sample \mathbf{x}_j and the test sample \mathbf{x} . In fact, our method can be viewed as a way for kernel combination which combines multiple kernels into one kernel.

It is worth noting that our formulation of multiple-kernel SVM is similar, but different from, the existing multi-kernel learning methods (Hinrichs et al., 2009b; Lanckriet et al., 2004; Wang et al., 2008). One key difference is that we do not jointly optimize the weights β_m s together with other SVM parameters (e.g., α) in an iterative way. Instead, we constrain $\sum_m \beta_m = 1$ and use a coarse-grid search through cross-validation on the training samples to find the optimal values. After we obtain the values of β_m s, we use them to combine multiple kernels into a mixed kernel, and then perform the standard SVM using the mixed kernel. The main advantage of our method is that it can be conveniently solved using the conventional SVM solvers, e.g., LIBSVM (Chang and Lin, 2001).

As explained above, this kernel combination method can provide a convenient and effective way for fusing various data from different modalities. In our case, we focus on multimodal classification using three modalities, i.e., MRI, PET, and CSF biomarkers. Figure 1 gives a schematic illustration of our multimodal data fusion and classification pipeline.

Validation

To evaluate the performance of different classification methods, we use 10-fold cross-validation strategy to compute the classification accuracy (for measuring the proportion of subjects correctly classified among the whole population), as well as the sensitivity (i.e., the proportion of AD or MCI patients correctly classified) and the specificity (i.e., the proportion of healthy controls correctly classified). Specifically, the whole set of subject samples are equally partitioned into 10 subsets, and each time the subject samples within one subset are successively selected as the testing samples and all remaining subject samples

in the other 9 subsets are used for training the multiple-kernel classifier. This process is repeated for 10 times independently to avoid any bias introduced by randomly partitioning dataset in the cross-validation. The SVM classifier is 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 are learned based on the training samples, through a grid search using the range from 0 to 1 at a step size of 0.1. Specifically, in each fold of the 10-fold cross-validation, we perform another 10-fold cross-validation on the training samples to determine the optimal values for the weights. Also, for each feature f_i in the training samples, a common feature normalization scheme is adopted,

i.e., $f_i = (f_i - \bar{f}_i) / \sigma_i$, where \bar{f}_i and σ_i are respectively the mean and standard deviation of the i -th feature across all training samples. The estimated \bar{f}_i and σ_i will be used to normalize the corresponding feature of each test sample.

Results

Multimodal classification based on MRI, PET, and CSF

We first test the performance of our multimodal classification method in identification of AD (or MCI) from healthy controls, based on MRI, PET, and CSF biomarkers of 202 baseline subjects in ADNI. Table 2 shows the classification rate of our multimodal classification method, compared with the methods using each individual modality only. Note that Table 2 shows only the averaged results of 10 independent experiments, along with the minimal and maximal values given in brackets; and the detailed results can be found in the supplemental Figs. 8-9 for each experiment. Besides, Fig. 2 further plots the corresponding ROC curves of different classification methods for AD or MCI, respectively. As we can see from Table 2 and Fig. 2, the combined measurements of MRI, PET, and CSF consistently achieve more accurate discrimination between AD (or MCI) patients and healthy controls. Specifically, for classifying AD from healthy controls, our multimodal classification method can achieve a classification accuracy of 93.2%, a sensitivity of 93%, 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 healthy controls, our multimodal classification method achieve a classification accuracy of 76.4%, a sensitivity of 81.8%, and a specificity of 66%, while the best accuracy on individual modality is only 72% (when using MRI). In addition, the area under the ROC curve (AUC) is 0.976 and 0.809 for AD classification and MCI classification respectively with our multimodal classification method (see Fig. 2), while the best AUC on individual modality is 0.938 (when using PET) for AD classification, and 0.762 (when using PET) for MCI classification.

Table 2 also indicates that, for AD classification, there are little differences among accuracy, sensitivity, and specificity of each classification method (totally 5 methods examined), while for MCI classification the differences are relatively large, e.g., relatively large sensitivity, but low specificity, for each method. This characteristic of possessing high sensitivity may be advantageous for diagnosis purpose, because the cost is different for misclassifying an MCI patient into a healthy control (with sensitivity reduced in this case) and misclassifying a healthy control into an MCI patient (with specificity reduced in this case), and the former cost is much higher than the latter. Inspired from this observation, we further divide the MCI cohort into MCI converters who converted to AD within 18 months and the MCI non-converters who had not convert to AD within 18 months, and then compute how many MCI converters and MCI non-converters are correctly classified as MCI. The results with our multimodal classification method reveal that the 91.5% MCI converters and 73.4% MCI non-converters are correctly classified. It's worth noting that in practice the cost of misclassifying MCI converters is usually much higher than that of misclassifying MCI non-

converters. Thus, this characteristic of possessing a higher classification rate for the MCI converters by our method is potentially very useful.

For comparison with other multimodal classification methods, we also perform the use of direct feature concatenation as a baseline method for multimodal AD (or MCI) classification. Specifically, for each subject, we first concatenate 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 perform SVM-based classification on all samples with a 10-fold cross-validation strategy as described above, and obtain the classification results in the bottom row of Table 2. As we can observe from Table 2, our kernel combination method consistently outperforms the baseline method on each performance measure.

Furthermore, in Table 3 we compared the proposed method with a recent method proposed in (Hinrichs et al., 2010). The latter used 114 ADNI subjects (48AD+66HC) for AD classification, and it reported both results of using only imaging modalities (MRI+PET) and all modalities (MRI+PET+CSF+APOE+ Cognitive scores), as included in Table 3. The proposed method uses a similar number of ADNI subjects, i.e., 103 subjects (51AD+52HC), with results given in Table 2. For comparison, we also include the proposed method's results in Table 3. As we can observe from Table 3, the proposed method is superior to Hinrichs et al.'s method in case of using only imaging modality (MRI+PET) or all modalities (MRI+PET+CSF). It's worth noting that, in (Hinrichs et al., 2010), both baseline and longitudinal data are used for MRI and PET modalities, while the proposed method uses only the baseline data. In the second case, even the additional APOE and cognitive scores were used in Hinrichs et al.'s method, our result is still better. These results further validate the efficacy of the proposed method for multimodal classification.

Comparison of different combination schemes

To investigate the effect of different combining weights, i.e., β_{MRI} , β_{CSF} , and β_{PET} , on the performance of our multimodal classification method, we test all of their possible values, ranging from 0 to 1 at a step size of 0.1, under the constraint of $\beta_{\text{MRI}}+\beta_{\text{CSF}}+\beta_{\text{PET}}=1$. Figures 3 and 4 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 $\beta_{\text{PET}}+\beta_{\text{CSF}}+\beta_{\text{MRI}}=1$. 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. 3 and 4, 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 substantially a large set of squares owning higher classification accuracy. Further observation indicates that the squares with higher accuracy mainly appear 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 2, Figs. 3 and 4 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.

Classification performance with respect to the number of selected ROI features

We have shown the effectiveness of our multiple-kernel combination method on using whole-brain ROI features (without feature selection) for AD or MCI classification. Here, we investigate how the performance of our multiple-kernel combination method changes with respect to the number of the selected ROI features. To this end, we first use a paired t-test, respectively, on MRI and PET data of training samples to choose the most discriminative brain regions or features for guiding AD or MCI classification (Gerardin et al., 2009). It's worth noting that the feature selection is performed using only the training samples, instead of all samples. Specifically, in each fold of the 10-fold cross-validations, we perform a t-test only on the training samples to select the most discriminative feature subset. Table 4 lists the top brain regions (or ROIs) detected from both MRI and PET data in MCI classification, and Figs. 5-6 show these top brain regions in the template space. Totally, 11 top brain regions, with corresponding p-values less than 0.002, are determined in MRI images. Notice that the top regions selected for AD classification are not listed, since the number is too large. As shown in Table 4 and Figs. 5-6, most of the selected top regions, e.g., hippocampal, amygdale, entorhinal cortex, uncus, temporal pole and parahippocampal regions, are known to be related to the AD by many studies using group comparison methods (Chetelat et al., 2002; Convit et al., 2000; Fox and Schott, 2004; Jack et al., 1999; Misra et al., 2009). For example, hippocampus is a structure highly related to the memory, which is always affected in the AD.

Then, we test the classification performances of different methods with respect to the different number of brain regions selected for AD (or MCI) classification, with results shown in Fig. 7. As we can see from Fig. 7, for both AD classification and MCI classification, our multimodal classification method (using all MRI, PET, and CSF) achieves consistent improvement over those using only one individual modality, for any number of brain regions selected. Moreover, compared with individual-modality based methods, our multimodal classification method is more robust to the number of brain regions used for classification. For example, Fig. 7 shows that, even only one brain region is selected for MRI and PET images, our multimodal classification method can still achieve a reasonable classification accuracy, compared to the individual-modality based classification methods. Another interesting observation from Fig. 7 is that more brain regions are needed for achieving higher accuracy for MCI classification than AD classification. This indicates that, with the progress of disease, more atrophies are produced in AD, thus a small number of brain regions with relatively large atrophies is sufficient for successful classification of AD.

Discussion

In this paper, we have proposed a new multimodal data fusion and classification method to automatically discriminate patients with AD (or MCI) from healthy controls, using a kernel combination method. This kernel combination method can be naturally embedded into the conventional SVM and solved efficiently. The results on 202 baseline subjects from ADNI show that our multimodal classification method can consistently and substantially improve the classification performance of the individual-modality based classification methods. Specifically, our method can achieve a high accuracy (93.2%) for AD classification, a relatively high sensitivity (81.8%) for MCI classification, and especially a high sensitivity (91.5%) for classification of MCI converters.

Multimodal data fusion and classification

A lot of studies have shown that biomarkers from different modalities may contain complementary information for diagnosis of AD (Apostolova et al., 2010; de Leon et al., 2007; Fjell et al., 2010; Foster et al., 2007; Landau et al., 2010; Walhovd et al., 2010b).

Recently, several works on combining different modalities of biomarkers have been reported (Bouwman et al., 2007a; Chetelat et al., 2005; Fan et al., 2008b; Fellgiebel et al., 2007; Geroldi et al., 2006; Vemuri et al., 2009; Visser et al., 2002; Walhovd et al., 2010a). A common practice in these works is the concatenation of all features (from different modalities) into a longer feature vector. However, this may be not enough for effective combination of features from different modalities. In this paper, we provide an alternative way by using 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 type of data cannot be directly concatenated; 2) it offers more flexibility by using different weights on biomarkers of different modalities. For instance, we cannot directly concatenate data represented by strings or graphs with numeric data while we can possibly construct separate kernels for string, graphs and numeric data respectively and then fuse them by kernel combination. In our case, since MRI, PET, and CSF are different types of features, the kernel combination provides us a better way to integrate them for guiding the classification.

It's worth noting that the kernel combination method has been successfully applied to many other fields, i.e., protein function prediction (Lanckriet et al., 2004), cancer diagnosis (Yu et al., 2010), and gene prioritization (De Bie et al., 2007). Recently, several researches have started to use this powerful kernel combination method for AD study (Hinrichs et al., 2009b; Ye et al., 2008). Specifically, in (Ye et al., 2008), MRI and APOE data as well as the age and sex information were combined using the existing multiple-kernel learning method. In (Hinrichs et al., 2009b), MRI and PET data were combined also using the same multiple-kernel learning method. However, both studies aimed only for AD classification, while in this paper we studied for both AD classification and MCI classification. The latter is actually more important than the former for early detection and treatment of AD. More importantly, we combine not only MRI and PET, but also CSF, which was rarely investigated before in the multiple-kernel combination study. Our experimental result shows that each modality (MRI, PET, and CSF) is indispensable for achieving good combination and classification. Also, we use more advanced feature extraction method with atlas warping, compared to those in (Hinrichs et al., 2009b; Ye et al., 2008). Thus, we can achieve much better performance compared to those reported in (Hinrichs et al., 2009b; Ye et al., 2008). Even for their new method using baseline MRI, PET, CSF, and additional longitudinal MRI and PET data, biological measures, and cognitive scores (Hinrichs et al., 2010), its performance is still inferior to our method using only baseline MRI, PET and CSF, as shown in Table 3.

Diversity of individual modalities in classification

As mentioned earlier, a lot of studies have indicated that different modalities contain complementary information for discrimination. Here, we quantitatively measure the discrimination similarity and diversity between any two different modalities, i.e., MRI vs CSF, MRI vs PET, and CSF vs PET, by comparing their individual classification results. Both Jaccard similarity coefficient and Kappa index are used to measure the similarities and diversities, respectively. Small values on both indexes imply a low similarity and a high diversity on the two modalities. For AD classification, the averaged similarities (*diversities*) over 10-fold cross-validation are 0.75 (0.53), 0.80 (0.62), and 0.74 (0.49) for MRI vs CSF, MRI vs PET, and CSF vs PET, respectively. On the other hand, for MCI classification, the averaged similarities (*diversities*) are 0.65 (0.33), 0.67 (0.38), and 0.63 (0.28), respectively. These results indicate that CSF and PET have the highest complementary information, while MRI and PET have the highest similar information for classification.

Data fusion vs ensemble

In this paper, we combine data from different modalities using kernel combination, which first combines multiple kernel matrices from different modalities into a single kernel matrix and then trains a single SVM model from the combined kernel matrix. Interestingly, we can also combine results from multiple modalities at classification stage. That is, we first train multiple SVM models on multiple kernel matrices from different modalities. Then, for a new testing sample, each of these models will have a predication on it, and finally we aggregate all predictions to get the final decision on the new testing sample. This technique is also called ensemble learning, which has been a very popular learning method for decades in the machine learning community (Tan and Gilbert, 2003).

We have compared our kernel combination method with the ensemble learning method for AD (or MCI) classification. Specifically, the ensemble learning method trains 3 SVM classifiers from MRI, PET, and CSF, respectively; and then the majority voting is used to get the final class labels for each new testing sample. The ensemble learning method obtains a classification accuracy of 91.8% for AD classification, and an accuracy of 75.6% for MCI classification, which are slightly inferior to the corresponding classification numbers achieved by our kernel combination method. These results indicate the effectiveness of the ensemble learning method as a useful and general way in improving classification accuracy of individual modalities. It may be even more interesting to investigate adding the mixed kernel from kernel combination into the ensemble or just ensembling different mixed kernels with different weights. However, the full investigation on this topic is beyond the focus of this paper. On the other hand, it is worth noting the disadvantage of the ensemble learning, i.e., the difficulty in interpreting the model since multiple models are used in the ensemble learning. This issue may limit its use in some medical applications where in addition to the accuracy, interpretability is also concerned and important.

Effect of feature selection

We test the kernel combination method on two cases, i.e. without and with feature selection. It is worth noting that the main concern of using feature selection in the current study is to validate the effectiveness of the kernel combination on the selected brain regions. Therefore, we adopt a simple feature selection method based on t-test statistics, which has been widely used in the neuroimaging analysis. Figure 7 shows that even a simple feature selection method can potentially select effective features (or regions) for achieving higher classification accuracy than the original methods using all features. We expect that the use of more advanced feature selection methods in the future can lead to further improvement for our multimodal classification.

On the other hand, in the current study we adopt a linear SVM as the classifier, which intrinsically uses a feature weighting mechanism, i.e., the absolute values of components in the normal vector of SVM's hyperplane can be regarded as weights on features (Kloppel et al., 2008). In this way, we can rank the features according to their averaged SVM weights. We find that the top-ranked features are partially identical with those top features obtained from a separate feature selection method we used. For example, among the top-ranked eleven features selected (according to SVM weights) for MCI classification on MRI modality, six features, namely, 'amygdala right', 'hippocampal formation left', 'hippocampal formation right', 'entorhinal cortex left', 'temporal pole left', and 'parahippocampal gyrus left', are identical to those selected by the t-test statistics as shown in Table 4. Notice that these six brain regions are known to be related to AD and MCI by many studies in the literature (Chetelat et al., 2002; Convit et al., 2000; Fox and Schott, 2004; Jack et al., 1999; Misra et al., 2009).

Limitations

While aiming to develop a multimodal diagnostic tool, the current study is limited by at least two factors. First, besides MRI, PET, and CSF, there are also other modalities of data, i.e., APOE. However, since not every subject has data on all modalities and the number of subjects with all modalities available is too small for reasonable classification, the current study does not consider APOE for multimodal classification. Second, in the current study, we investigate only the classification between one stage of dementia (either MCI or AD) and healthy controls, and do not test the ability of the classifier to simultaneously discriminate multiple stages of dementia, i.e., multi-class classification of AD, MCI, and healthy controls. Although the conversion from binary-class classification to multi-class classification seems straightforward, with many multi-class classification methods available (Duda et al., 2001), there may be some problem and this will be our future work.

Conclusion

This study proposes a new multimodal data fusion and classification method based on kernel combination for AD and MCI. Compared with the conventional direct feature concatenation method, our method provides a unified way to combine heterogeneous data, particularly for the case where different types of data cannot be directly concatenated. Moreover, our method offers more flexibility by using different weights for different data modalities. The results on 202 baseline subjects of ADNI show that our multimodal classification method achieves a high accuracy for AD classification and an encouraging accuracy for MCI classification.

The current study only considers the baseline data of the subjects in ADNI. In the future, we will use both baseline and longitudinal data to predict the conversion from MCI to AD by finding the spatiotemporal pattern of brain atrophy in multiple modalities. Moreover, we will involve using more modalities of data (i.e., APOE) into our current multimodal classification method. To overcome the limitation of the possible small number of subjects available for training and testing classifier as discussed earlier, we will seek more advanced methods in machine learning which can use missing data for classification, i.e., semi-supervised classification. We expect that, by using more samples (with both complete and missing modality information), the semi-supervised method will improve the classification performance further.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

- Apostolova LG, Hwang KS, Andrawis JP, Green AE, Babakchianian S, Morra JH, Cummings JL, Toga AW, Trojanowski JQ, Shaw LM, Jack CR Jr, Petersen RC, Aisen PS, Jagust WJ, Koeppe RA, Mathis CA, Weiner MW, Thompson PM. 3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects. *Neurobiol Aging* 2010;31:1284–1303. [PubMed: 20538372]
- Bouwman FH, Schoonenboom SN, van der Flier WM, van Elk EJ, Kok A, Barkhof F, Blankenstein MA, Scheltens P. CSF biomarkers and medial temporal lobe atrophy predict dementia in mild cognitive impairment. *Neurobiol Aging* 2007a;28:1070–1074. [PubMed: 16782233]
- Bouwman FH, van der Flier WM, Schoonenboom NS, van Elk EJ, Kok A, Rijmen F, Blankenstein MA, Scheltens P. Longitudinal changes of CSF biomarkers in memory clinic patients. *Neurology* 2007b;69:1006–1011. [PubMed: 17785669]
- Chang CC, Lin CJ. LIBSVM: a library for support vector machines. 2001
- Chetelat G, Desgranges B, de la Sayette V, Viader F, Eustache F, Baron J-C. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport* 2002;13:1939–1943. [PubMed: 12395096]
- Chetelat G, Desgranges B, de la Sayette V, Viader F, Eustache F, Baron JC. Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology* 2003;60:1374–1377. [PubMed: 12707450]
- Chetelat G, Eustache F, Viader F, De La Sayette V, Pelerin A, Mezenge F, Hannequin D, Dupuy B, Baron JC, Desgranges B. FDG-PET measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment. *Neurocase* 2005;11:14–25. [PubMed: 15804920]
- Convit A, de Asis J, de Leon MJ, Tarshish CY, De Santi S, Rusinek H. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiology of Aging* 2000;21:19–26. [PubMed: 10794844]
- 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*. 2010 in press.
- 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 in press.
- De Bie T, Tranchevent LC, van Oeffelen LM, Moreau Y. Kernel-based data fusion for gene prioritization. *Bioinformatics* 2007;23:i125–132. [PubMed: 17646288]
- de Leon MJ, Mosconi L, Li J, De Santi S, Yao Y, Tsui WH, Pirraglia E, Rich K, Javier E, Brys M, Glodzik L, Switalski R, Saint Louis LA, Pratico D. Longitudinal CSF isoprostane and MRI atrophy in the progression to AD. *J Neurol* 2007;254:1666–1675. [PubMed: 17994313]
- De Santi S, de Leon MJ, Rusinek H, Convit A, Tarshish CY, Roche A, Tsui WH, Kandil E, Boppana M, Daisley K, Wang GJ, Schlyer D, Fowler J. Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiology of Aging* 2001;22:529–539. [PubMed: 11445252]
- Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL, Fischl B. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 2009;132:2048–2057. [PubMed: 19460794]
- Diehl J, Grimmer T, Drzezga A, Riemenschneider M, Forstl H, Kurz A. Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. *Neurobiology of Aging* 2004;25:1051–1056. [PubMed: 15212830]
- Drzezga A, Lautenschlager N, Siebner H, Riemenschneider M, Willoch F, Minoshima S, Schwaiger M, Kurz A. Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study. *Eur J Nucl Med Mol Imaging* 2003;30:1104–1113. [PubMed: 12764551]
- Du AT, Schuff N, Kramer JH, Rosen HJ, Gorno-Tempini ML, Rankin K, Miller BL, Weiner MW. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* 2007;130:1159–1166. [PubMed: 17353226]
- Duda, RO.; Hart, PE.; Stork, DG. *Pattern Classification*. John Wiley and Sons, Inc.; 2001.

- Fan Y, Batmanghelich N, Clark CM, Davatzikos C, Initiative, t.A.s.D.N. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 2008a;39:1731–1743. [PubMed: 18053747]
- Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage* 2008b;41:277–285. [PubMed: 18400519]
- Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: Classification Of Morphological Patterns using Adaptive Regional Elements. *IEEE Transactions on Medical Imaging* 2007;26:93–105. [PubMed: 17243588]
- Fellgiebel A, Scheurich A, Bartenstein P, Muller MJ. FDG-PET and CSF phospho-tau for prediction of cognitive decline in mild cognitive impairment. *Psychiatry Res* 2007;155:167–171. [PubMed: 17531450]
- Fjell AM, Walhovd KB, Fennema-Notestine C, McEvoy LK, Hagler DJ, Holland D, Brewer JB, Dale AM. CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. *J Neurosci* 2010;30:2088–2101. [PubMed: 20147537]
- Foster NL, Heidebrink JL, Clark CM, Jagust WJ, Arnold SE, Barbas NR, DeCarli CS, Turner RS, Koeppe RA, Higdon R, Minoshima S. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. *Brain* 2007;130:2616–2635. [PubMed: 17704526]
- Fox N, Schott J. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *Lancet* 2004;363:392–394. [PubMed: 15074306]
- Gerardin E, Chetelat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, Niethammer M, Dubois B, Lehericy S, Garnero L, Eustache F, Colliot O. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage* 2009;47:1476–1486. [PubMed: 19463957]
- Geroldi C, Rossi R, Calvagna C, Testa C, Bresciani L, Binetti G, Zanetti O, Frisoni GB. Medial temporal atrophy but not memory deficit predicts progression to dementia in patients with mild cognitive impairment. *Journal of Neurology Neurosurgery and Psychiatry* 2006;77:1219–1222.
- Higdon R, Foster NL, Koeppe RA, DeCarli CS, Jagust WJ, Clark CM, Barbas NR, Arnold SE, Turner RS, Heidebrink JL, Minoshima S. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging. *Statistics in Medicine* 2004;23:315–326. [PubMed: 14716732]
- 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* 2009a;48:138–149. [PubMed: 19481161]
- Hinrichs C, Singh V, Xu G, Johnson S. MKL for robust multi-modality AD classification. *Med Image Comput Assist Interv* 2009b;12:786–794. [PubMed: 20426183]
- Hinrichs C, Singh V, Xu G, Johnson S. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage*. 2010 in press.
- Jack CR Jr. Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 2010;9:119–128. [PubMed: 20083042]
- Jack CR, Petersen RC, X. YC, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Waring SC, Tangalos E, Kokmen E. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology* 1999;52:1397–1403. [PubMed: 10227624]
- Ji Y, Permanne B, Sigurdsson EM, Holtzman DM, Wisniewski T. Amyloid b40/42 clearance across the blood-brain barrier following intra-ventricular injections in wild-type, apoE knock-out and human apoE3 or E4 expressing transgenic mice. *Journal of Alzheimer's Disease* 2001;3:23–30.
- Kabani N, MacDonald D, Holmes CJ, Evans A. A 3D atlas of the human brain. *Neuroimage* 1998;7:S717.
- Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr. Ashburner J, Frackowiak RS. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008;131:681–689. [PubMed: 18202106]

- Lanckriet GR, Deng M, Cristianini N, Jordan MI, Noble WS. Kernel-based data fusion and its application to protein function prediction in yeast. *Pac Symp Biocomput* 2004;300–311. [PubMed: 14992512]
- Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack CR Jr. Weiner MW, Jagust WJ. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 2010;75:230–238. [PubMed: 20592257]
- 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]
- Leuchter JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, Evans AC. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of Aging* 2008;29:23–30. [PubMed: 17097767]
- 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]
- Mattsson N, Zetterberg H, Hansson O, Andreassen N, Parnetti L, Jonsson M, Herukka SK, van der Flier WM, Blankenstein MA, Ewers M, Rich K, Kaiser E, Verbeek M, Tsolaki M, Mulugeta E, Rosen E, Aarsland D, Visser PJ, Schroder J, Marcusson J, de Leon M, Hampel H, Scheltens P, Pirtila T, Wallin A, Jonhagen ME, Minthon L, Winblad B, Blennow K. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *Jama* 2009;302:385–393. [PubMed: 19622817]
- McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ Jr. Holland D, Karow DS, Pung CJ, Brewer JB, Dale AM. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology* 2009;251:195–205. [PubMed: 19201945]
- Misra C, Fan Y, Davatzikos C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* 2009;44:1415–1422. [PubMed: 19027862]
- Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease. *Archives of Neurology* 2001;58:397–405. [PubMed: 11255443]
- Nestor PJ, Scheltens P, Hodges JR. Advances in the early detection of Alzheimer's disease. *Nat Med* 2004;10(Suppl):S34–41. [PubMed: 15298007]
- Oliveira PJ, Nitri 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 Alzheimers Dis* 2010;18:1263–1272. [PubMed: 20061613]
- 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]
- Ron B, Elizabeth J, Kathryn Z-G, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2007;3:186–191.
- Scholkopf, B.; Smola, A. *Learning with Kernels*. The MIT Press; 2002.
- Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahy RM. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage* 2001;13:856–876. [PubMed: 11304082]
- Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* 2009;65:403–413. [PubMed: 19296504]
- Shen D, Davatzikos C. HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging* 2002;21:1421–1439. [PubMed: 12575879]
- 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]

- Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002;17:143–155. [PubMed: 12391568]
- Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics* 2003;2:S75–83. [PubMed: 15130820]
- Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR Jr. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology* 2009;73:294–301. [PubMed: 19636049]
- Visser PJ, Verhey FRJ, Hofman PA, Scheltens P, Jolles J. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry* 2002;72:491–497.
- Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler DJ Jr, Jennings RG, Karow D, Dale AM. Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *AJNR Am J Neuroradiol* 2010a; 31:347–354. [PubMed: 20075088]
- Walhovd KB, Fjell AM, Dale AM, McEvoy LK, Brewer J, Karow DS, Salmon DP, Fennema-Notestine C. Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobiol Aging* 2010b;31:1107–1121. [PubMed: 18838195]
- Wang Z, Chen S, Sun T. MultiK-MHKS: a novel multiple kernel learning algorithm. *IEEE Trans Pattern Anal Mach Intell* 2008;30:348–353. [PubMed: 18084064]
- West M, Kawas C, Stewart W, Rudow G, Troncoso J. Hippocampal neurons in pre-clinical Alzheimer's disease. *Neurobiology of Aging* 2004;25:1205–1212. [PubMed: 15312966]
- Westman E, Simmons A, Zhang Y, Muehlboeck JS, Tunnard C, Liu Y, Collins L, Evans A, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Wahlund LO. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage*. 2010 in press.
- Ye J, Chen K, Wu T, Li J, Zhao Z, Patel R, Bae M, Janardan R, Liu H, Alexander G, Reiman EM. Heterogeneous data fusion for Alzheimer's disease study. *KDD'08*. 2008
- Yu S, Falck T, Daemen A, Tranchevent LC, Suykens JA, De Moor B, Moreau Y. L2-norm multiple kernel learning and its application to biomedical data fusion. *BMC Bioinformatics* 2010;11:309. [PubMed: 20529363]
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging* 2001;20:45–57. [PubMed: 11293691]

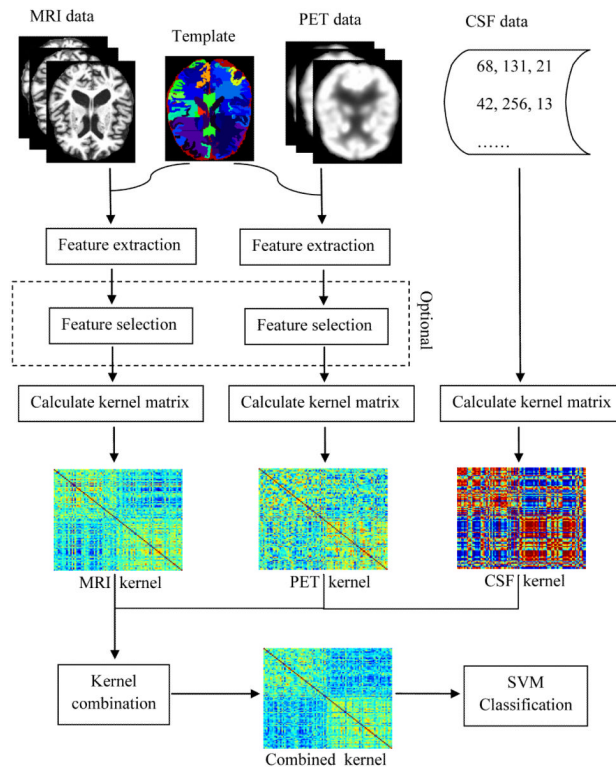


Fig. 1. Schematic illustration of multimodal data fusion and classification pipeline.

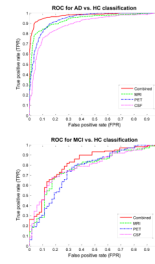


Fig. 2. ROC curves of different methods, for AD classification (top) and for MCI classification (bottom).

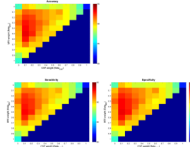


Fig. 3. AD Classification results with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint: $\beta_{\text{PET}} + \beta_{\text{CSF}} + \beta_{\text{MRI}} = 1$. Note that for each plot, the top left, top right, 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.

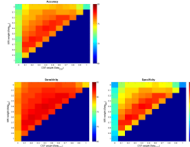


Fig. 4. MCI Classification with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint: $\beta_{PET} + \beta_{CSF} + \beta_{MRI} = 1$. Note that for each plot, the top left, top right, and bottom left squares denote the individual-modality based classification results using PET ($\beta_{PET}=1$), CSF ($\beta_{CSF}=1$) and MRI ($\beta_{MRI}=1$), respectively.

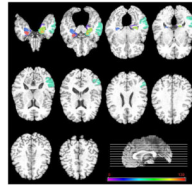


Fig. 5.
Top 11 brain regions selected for MCI classification detected from MRI. Brain regions are overlaid on the template image, and images are displayed in radiological convention.

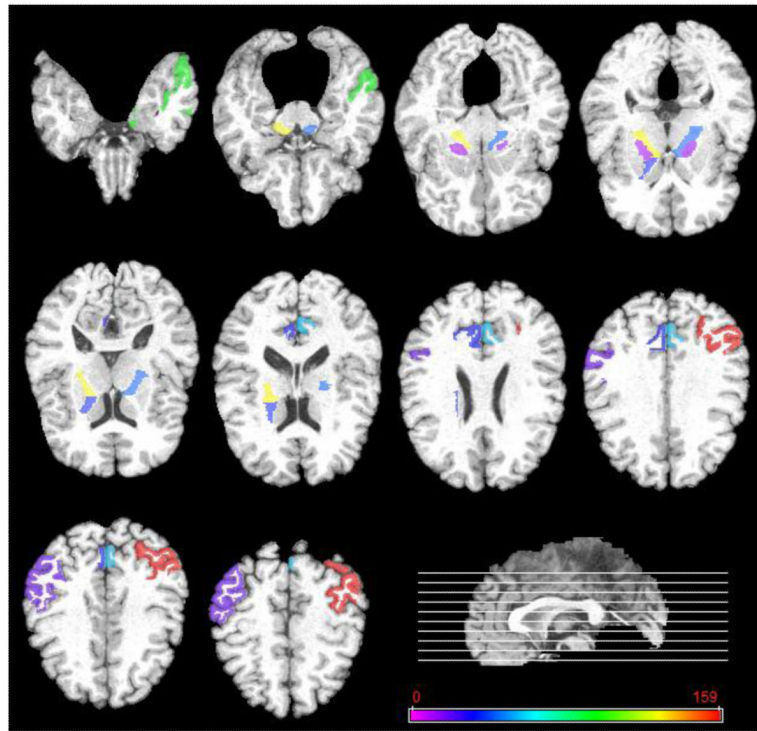


Fig. 6. Top 11 brain regions selected for MCI classification detected from PET. Brain regions are overlaid on the template image, and images are displayed in radiological convention.

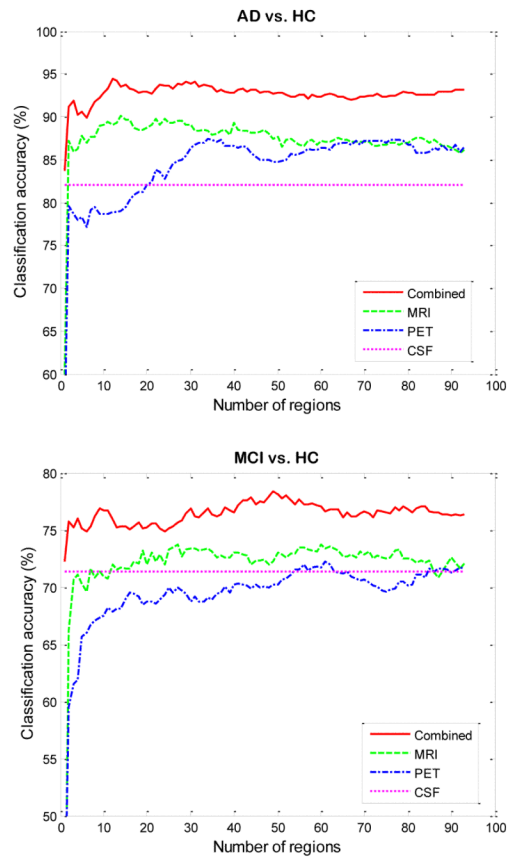


Fig. 7. Classification accuracy of four different methods, with respect to different number of regions selected for AD classification (top) and MCI classification (bottom).

Table 1

Subject information

	<u>AD (n=51; 18F/33M)</u>			<u>MCI (n=99; 32F/67M)</u>			<u>HC (n=52; 18F/34M)</u>		
	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	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

The numbers refer to baseline data. AD = Alzheimer's Disease, MCI = Mild Cognitive Impairment, HC = Healthy Control, MMSE = Mini-Mental State Examination, CDR = Clinical Dementia Rating

Table 2

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

Methods	AD vs. HC			MCI vs. HC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
MRI	86.2 (82.9-89.0)	86 (82.7-88.7)	86.3 (83.1-89.1)	72.0 (68.4-74.7)	78.5 (75.6-80.6)	59.6 (55.1-63.7)
CSF	82.1 (80-84.9)	81.9 (80-84.7)	82.3 (80-85.1)	71.4 (68.2-73.3)	78 (75.6-79.4)	58.8 (54.3-61.7)
PET	86.5 (82.9-90.5)	86.3 (82.7-90.3)	86.6 (83.1-90.6)	71.6 (67.4-74.7)	78.2 (75-80.6)	59.3 (52.9-63.7)
Combined	93.2 (89.0-96.5)	93 (88.7-96.3)	93.3 (89.1-96.6)	76.4 (73.5-79.7)	81.8 (79.4-84.4)	66.0 (62.6-70.3)
Baseline	91.5 (88.5-96.5)	91.4 (88.3-96.3)	91.6 (88.6-96.6)	74.5 (71.9-78.2)	80.4 (78.3-83.3)	63.3 (59.7-68.3)

AD = Alzheimer's Disease, MCI = Mild Cognitive Impairment, HC = healthy control, ACC = classification Accuracy, SEN = SENSitivity, SPE = SPEcificity

Table 3

Comparison of performance of different multimodal classification methods

Methods	Subjects	Modalities	ACC (%)	SEN (%)	SPE (%)
Hinrichs et al., 2010	48AD+66HC	<i>MRI+PET</i>	87.6	78.9	93.8
		<i>MRI+PET+CSF+APOE</i> +Cognitive scores	92.4	86.7	96.6
Proposed method	51AD+52HC	<i>MRI+PET</i>	90.6	90.5	90.7
		<i>MRI+PET+CSF</i>	93.2	93	93.3

AD = Alzheimer's Disease, HC = healthy control, ACC = classification Accuracy, SEN = SENSitivity, SPE = SPEificity

Table 4

Top 11 brain regions detected from MRI and PET modalities for MCI classification (ranked according to the p-values in the brackets)

	MRI	PET
1	amygdala right (p<0.0001)	angular gyrus left (p=0.0003)
2	hippocampal formation left (p<0.0001)	precuneus left (p=0.0005)
3	hippocampal formation right (p<0.0001)	precuneus right (p=0.0021)
4	uncus left (p<0.0001)	inferior temporal gyrus left (p=0.0146)
5	entorhinal cortex left (p=0.0001)	anterior limb of internal capsule right (p=0.0154)
6	amygdala left (p=0.0001)	angular gyrus right (p=0.0189)
7	middle temporal gyrus left (p=0.0001)	anterior limb of internal capsule left (p=0.0204)
8	temporal pole left (p=0.0004)	globus palladus left (p=0.021)
9	perirhinal cortex left (p=0.0004)	globus palladus right (p=0.0259)
10	uncus right (p=0.0006)	posterior limb of internal capsule right (p=0.0272)
11	parahippocampal gyrus left (p=0.0009)	entorhinal cortex left (p=0.0286)