Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging

Christos Davatzikos [a,*], Yong Fan [a], Xiaoying Wu [a],
Dinggang Shen [a], Susan M. Resnick [b]

[a] *Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, United States*
[b] *Laboratory of Personality and Cognition, National Institute on Aging, United States*

Abstract

We report evidence that computer-based high-dimensional pattern classification of magnetic resonance imaging (MRI) detects patterns of brain structure characterizing mild cognitive impairment (MCI), often a prodromal phase of Alzheimer's disease (AD). Ninety percent diagnostic accuracy was achieved, using cross-validation, for 30 participants in the Baltimore Longitudinal Study of Aging. Retrospective evaluation of serial scans obtained during prior years revealed gradual increases in structural abnormality for the MCI group, often before clinical symptoms, but slower increase for individuals remaining cognitively normal. Detecting complex patterns of brain abnormality in very early stages of cognitive impairment has pivotal importance for the detection and management of AD.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Prodromal Alzheimer's disease; MCI; Pattern recognition; MRI

1. Introduction

Prevalence of AD doubles every 5 years of life after age 60, with more than 4 million individuals affected in the US alone. AD is the most common dementing illness and a major public health issue of increasing importance as life expectancy increases. Although noninvasive approaches for antemortem diagnosis of AD are under development, definitive diagnosis of AD requires neuropathologic confirmation of the characteristic amyloid plaques and neurofibrillary tangles (Braak et al., 1998). New drugs under development will target different stages of disease pathophysiology, and efficacious AD treatments likely will require early initiation before irreversible brain tissue damage. Thus, a great deal of attention has been paid recently to the prodromal stage of AD, referred to as mild cognitive impairment (MCI), which includes individuals with memory problems who do not meet criteria for dementia. Although MCI definitions vary across studies (Petersen, 2003), MCI individuals convert to AD with rates of 6–15% annually (Petersen, 2003). Therefore, MCI individuals are a high risk group likely to benefit from effective treatments.

Structural magnetic resonance imaging (MRI) promises to aid diagnosis and treatment monitoring of MCI and AD, offering the potential for easily obtainable surrogate markers of diagnostic status and disease progression. Unlike relatively more advanced stages of MCI and AD, quantifying patterns of structural change during early stages of AD or during clinically normal stages is a major challenge. Brain atrophy in the early stages of AD may be relatively subtle and spatially distributed over many brain regions (Chetelat et al., 2002; Convit et al., 2000; Dickerson et al., 2001; Kaye et al., 1997; Killiany et al., 2000b), including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, anterior and posterior cingulate, and possibly other regions that have only recently been investigated (Medina et al., 2006). Furthermore, spatially heterogeneous patterns of atrophy have been found within the hippocampus, with regions known to correspond to the CA1 field presenting relatively more pro-

* Corresponding author at: Department of Radiology, University of Pennsylvania, 3600 Market Street, Suite 380, Philadelphia, PA 19104, USA.
Tel.: +1 215 349 8587; fax: +1 215 614 0266.
E-mail addresses: christos@rad.upenn.edu,
christos.davatzikos@uphs.upenn.edu (C. Davatzikos).

nounced atrophy (Frisoni et al., 2006; Wang et al., 2006). Patterns of atrophy associated with pathology are confounded by complex patterns of atrophy associated with normal aging (Resnick et al., 2000). Moreover, the error associated with structural measurements can vary throughout the brain, since some structures are more difficult to delineate, especially via computer algorithms, thereby rendering the measurement of certain brain regions more informative than others merely for methodological reasons (Bookstein, 2001). Therefore, powerful and sensitive statistical image analysis methods must be used to capture morphological characteristics that are different between normal aging and MCI, and to determine which are most informative, from a diagnostic perspective.

Most MRI studies in MCI and AD have relied on measurement of volumes of specific brain regions (Chetelat, 2003; Nestor et al., 2004), especially the hippocampus and the entorhinal cortex, which show histopathogical changes at early stages of AD (Braak et al., 1998). Computational neuroanatomy has also been used to evaluate voxel-by-voxel brain changes in healthy aging, MCI and AD (Ashburner et al., 2003). These studies have confirmed patterns of atrophy involving medial temporal lobe structures in MCI and AD. They have reinforced the value of MRI as a potential surrogate marker of disease at the *group-analysis* level, i.e. for examining overall differences between individuals with and without pathology. However, their *diagnostic* value is limited, especially at early stages of brain pathology, since their sensitivity and specificity are not sufficient for prediction of the status of a given individual.

Herein we report results from a longitudinal study that provide strong evidence that there is a subtle and spatially distributed pattern of brain structure that is characteristic of MCI, and which often begins developing prior to the recognition of cognitive deficits. Moreover, this pattern can be detected with high sensitivity and specificity using a high-dimensional image analysis and pattern classification method that examines spatial patterns of brain atrophy in their entirety, instead of applying separate region-by-region evaluations. Therefore, detection of this structural pattern can lead to very early diagnosis of prodromal AD. This study adds to mounting evidence in the literature for the importance of pattern classification methods in detecting subtle and complex structural and functional patterns (Davatzikos et al., 2005a,b; Golland et al., 2002; Liu et al., 2004).

2. Materials and methods

2.1. Subjects

MRI scans from 30 elderly individuals were obtained annually as part of the Baltimore Longitudinal Study of Aging neuroimaging substudy (Resnick et al., 2000). At initial enrollment, all individuals were free of dementia and other central nervous system disorders, severe cardiovascular disease, and metastatic cancer (detailed in (Resnick et al., 2000)). Screening of mental status by the blessed-information-memory-concentration (BIMC) test was performed at each annual visit in conjunction with a comprehensive neuropsychological assessment. Subject and informant based assessments with the clinical dementia rating (CDR) (Morris et al., 1989) scale were administered by certified examiners to participants in the BLSA autopsy study annually (about 50% of participants) and to remaining participants scoring 3 or more BIMC errors. After as many as 9 annual neuroimaging and cognitive assessments, 20 participants with MCI were identified from a sample of 155 neuroimaging study participants who completed MRI studies. These 20 participants were characterized as MCI, based on CDR scores of 0.5 and/or consensus diagnosis indicating memory impairment that does not meet criteria for dementia. Of these individuals, 15 were eligible for inclusion in the current study. Five participants who developed cognitive impairment over the course of the study were excluded from these analyses due to documentation of other pathological processes (e.g., clinical stroke (1), brain trauma (1), heavy alcohol use (1), post-surgical confusion (1), absence of Alzheimer's pathology at autopsy (1)). A control sample of 15 individuals who remained unimpaired (CDR = 0), matched for age, sex, and follow-up interval, was identified from the remaining 135 participants. Subject characteristics are shown in Table 1. It is important to emphasize that the MCI participants in this study are identified within the context of prospective longitudinal follow-ups and typically represent relatively mild cases of cognitive impairment in contrast to those followed in other studies who typically present with memory complaints (e.g. Grundman et al., 2004). None had CDR total scores greater than 0.5 for any of the visits used in these analyses, and the mean (S.D.) of the sum of the individual CDR box scores was 1.2 (0.9) for the most recent visit included in the analyses. Therefore, in this group AD pathology is likely to be at a relatively early stage. To date, 10 of the 15 MCI individuals have been assigned diagnoses of Alzheimer disease at subsequent follow-up visits, verifying that we are including individuals who are progressing as well as those who may be more stable.

2.2. Imaging protocol

MR acquisition procedures are detailed in (Resnick et al., 2000). MR scanning was performed on a GE

Table 1
Characteristics of the participants in this MCI study

Group	MCI	Normal
No. of subjects	15	15
Sex no. of males	10	10
No. of left-handed	0	1
Years of education, mean (S.D.)	15.80 (3.73)	16.73 (3.22)
Baseline age, mean (S.D.)	76.92 (7.28)	75.21 (6.85)
Age at last visit, mean (S.D.)	82.40 (6.59)	81.76 (6.57)
Follow-up interval, mean (S.D.)	5.49 (2.42)	6.55 (2.51)
MMSE at last visit, mean (S.D.)	25.80 (2.96)	29.00 (1.41)

Signa 1.5 Tesla scanner. The current results are based on a high-resolution volumetric "spoiled grass" (SPGR) series (axial acquisition, TR = 35, TE = 5, flip angle = 45, FOV = 24, matrix = 256 × 256, NEX = 1, voxel dimensions of 0.94 mm × 0.94 mm × 1.5 mm slice thickness).

2.3. Image analysis

Images were pre-processed using the methods described in (Goldszal et al., 1998), which resulted in segmentations into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). To compare structural patterns across individuals, we spatially transformed each segmented image into a common coordinate system, often called stereotaxic space. A mass-preserving framework was adopted (Shen and Davatzikos, 2003) to ensure that the volumes of brain tissue were preserved during the transformation and to provide tissue density maps of GM, WM and CSF for each individual that reflected the spatial distribution of these tissue volumes. For example, relatively lower GM tissue density in the hippocampus would be indicative of hippocampal atrophy.

2.4. Pattern analysis and classification

Patterns of the spatial distribution of GM, WM and CSF volumes were then examined via a pattern classification technique (Fan et al., 2005), and patterns specific to MCI were determined. In particular, regions in which the tissue density correlated well with the clinical variable (MCI = 1, normal = −1) were first identified, via a voxel-by-voxel calculation of the Pearson correlation coefficient. In order to render this calculation robust to outliers, a leave-one-out procedure was applied, i.e. given n training samples, the correlation coefficients were calculated n times, each time leaving one of the scans out. The minimum value was then retained, representing the worst case scenario. This approach allowed us to subsequently construct spatial patterns from brain regions that were not only good discriminators between normal and MCI groups, but also were robust. Additional robustness was achieved by examining the spatial consistency between a voxel and its spatial neighborhood, and retaining only the brain regions that displayed both robust correlation with clinical status and high spatial consistency. A watershed-based clustering method was then used to determine brain regions whose volumes had good discriminant features. Finally, a recursive feature elimination procedure was used to find a minimal set of features to be fed to the classifier. Because the predictive power varies somewhat as a function of the number of brain regions (features), we estimated predictive power by averaging the abnormality scores (see below) obtained from all classifiers built for cluster numbers ranging from 12 to 45 (where the predictive power reaches a plateau). Additional details of the feature construction and selection procedures can be found in (Fan et al., 2005).

This analysis was cross-sectional, in that it was based solely on the tissue density maps obtained for the most recent

MRI assessment for each individual. For participants who developed dementia over the course of the study, the most recent MRI prior to the diagnosis was used, with a mean (S.D.) interval of 1.9 (1.9) years between most recent scan and diagnosis. Volumetric measurements from these brain regions were then used to build a classifier (Lao et al., 2004; Vapnik, 1998), which produced an abnormality score: positive values indicate a structural pattern resembling MCI, whereas negative values indicate brain structure in unimpaired individuals. A value of 0 would indicate a structural profile that is in-between normal and abnormal. Leave-one-out cross-validation was used to test the predictive power of this analysis on new datasets (datasets not involved in the selection of optimal brain clusters and training of the classifier) and construct receiver operating characteristic (ROC) curves that summarize predictive value. In this analysis, the scan of one participant was put aside, and the classifier was constructed from the most recent scans of all other individuals. Thus, the individual being classified was not included in the training data set for development of the classifier. This classifier was then applied to all available scans of the left-out individual. In this way, the temporal evolution of these spatial patterns of brain abnormality was measured during earlier years for each individual.

### 2.5. MCI-specific spatial pattern of brain structure

In order to form an image of the brain regions that constitute a pattern of brain tissue distribution that is characteristic of MCI, we followed our earlier work (Lao et al., 2004), and formed a spatial map of brain regions whose volumes change as one follows the path of fastest change in the abnormality score from positive to negative. These regions jointly form a pattern that optimally characterizes the differences between MCI and healthy controls, from the perspective of classification. Moreover, a value from 0 to 1 is determined for each region, reflecting its relative importance in classification.

The path of fastest change was constructed by taking each support vector, i.e. each sample that was close to the interface between MCI and controls, and following the gradient of the SVM's decision function. This gradient is known to provide the direction of fastest change. In the context of our approach, when following the gradient direction, one moves from an MCI sample to the opposite side, i.e. to the realm of features corresponding to normal controls, thereby generating a spatial map of the regions whose volumes change when one "makes an MCI brain look like healthy". This procedure was repeated for each support vector, and an average spatial map was generated for display purposes.
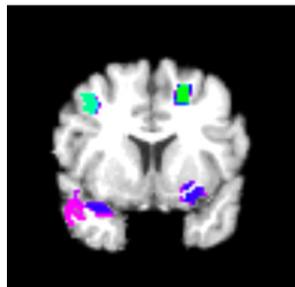
3. Results

3.1. Spatial pattern of MCI-specific abnormalities

The spatial map of brain regions that was formed as described in Section 2 is shown in Fig. 1a. This set of
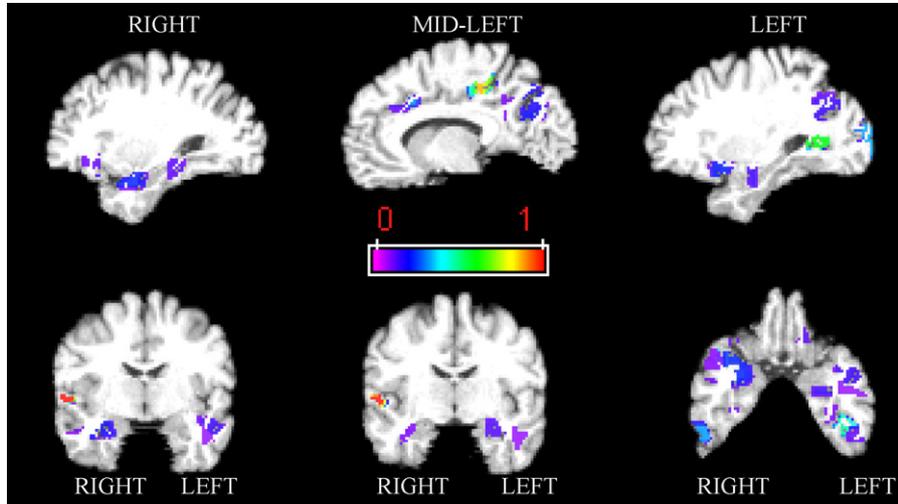
Fig. 1. Representative cross-sections highlighting the brain regions that collectively form a spatial pattern of brain atrophy that is highly indicative of MCI. The color-coding shows relative importance of a brain region for classification. The regions in (a) were located in areas known to be implicated in AD. The superior and middle frontal WM clusters in (b) were unexpected and must be further investigated via methods particularly suitable for WM structures, such as diffusion tensor imaging.

regions forms a structural network, which according to our classification approach, carries the most distinctive characteristics of MCI relative to unimpaired individuals. This map highlights several regions including the lateral and inferior aspects of both hippocampi, which is where the CA1 field of the hippocampus is located, bilateral superior, middle and inferior temporal gyri (GM), bilateral orbitofrontal GM, left fusiform gyrus (GM), right collateral sulcus, and cingulate, especially posteriorly. We also found clusters of reduced WM volumes in the inferior temporal gyri as well as middle and superior frontal gyri (see Fig. 1b).

3.2. Predictive accuracy

Based on the pattern of spatial distribution of brain tissue in the regions of Fig. 1, individuals with MCI were distinguished initially with 100% accuracy from those without cognitive impairment, thereby demonstrating full separation between MCI and cognitively normal individuals, using this nonlinear classifier. Next, predictive power was determined via the leave-one-out cross-validation (Section 2) to be 90%; this is an estimate of classification accuracy of a new individ-

ual's scan and therefore of direct diagnostic relevance. Fig. 2 shows the ROC curve of the leave-one-out analysis, indicating very good predictive power and sensitivity/specificity trade-off.
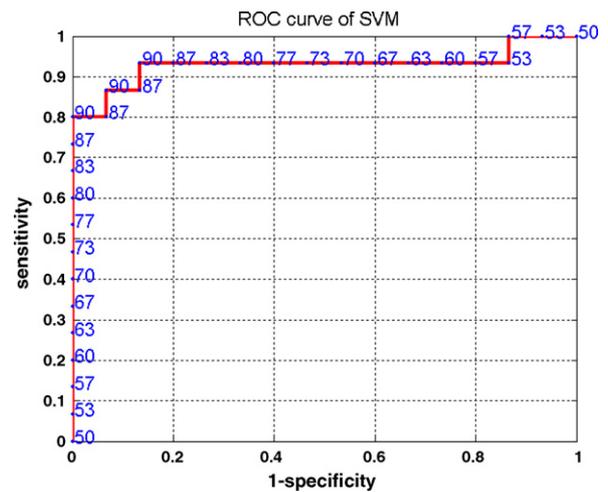

Fig. 2. ROC plots showing the trade-off between specificity and sensitivity in MRI-based detection of MCI. For each point on the ROC, the overall accuracy is shown in blue numbers (0.90 = 90%).

3.3. Retrospective evaluation of abnormality

As described earlier, the classifier was determined from the latest scans of all participants and did not include prior longitudinal images. Next, the classifier derived from the leave-one-out analysis of the most recent scans was retrospectively applied to all available earlier scans for each participant, yielding an abnormality score for each individual scan for each participant. The longitudinal evolutions of the abnormality scores for all available scans of each participant are shown in Fig. 3a. As described above, the values for each individual were based on cross-validation, i.e. all scans of a given individual were omitted during the construction of the classifier for that individual based on the most recent scans of all other participants. Fig. 3a shows increases over time in the structural abnormality scores from normal levels (negative) to MCI-like levels (positive). Mixed
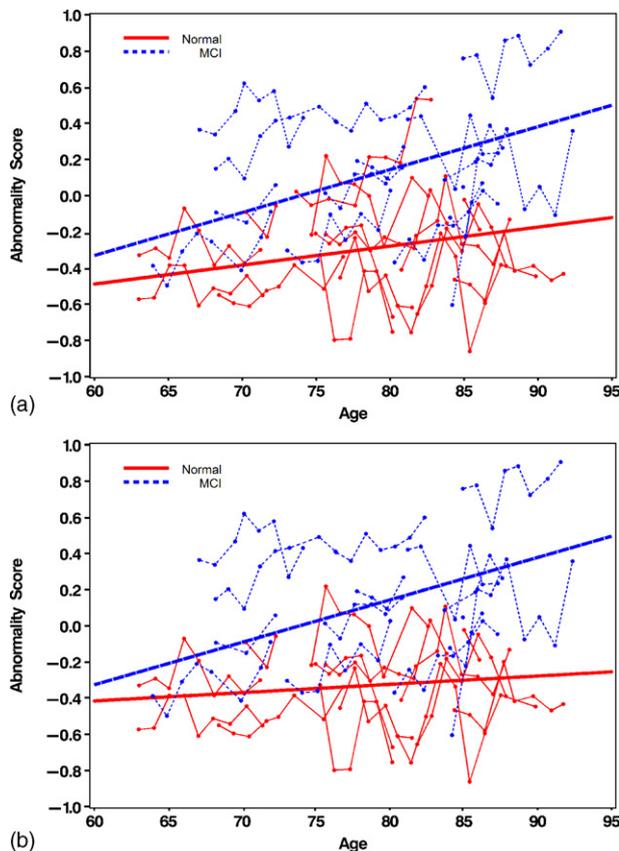

(a)

(b)

Fig. 3. Longitudinal change in the degree of structural abnormality measured by applying the pattern classification technique retrospectively to scans prior to the year in which the scans of healthy individuals and MCI patients were analyzed. Solid lines show estimates of the mean longitudinal changes for the two groups, derived from a mixed effects regression model. The faster increase of the abnormality score in the individuals who develop MCI reflects an important finding of this study: structural changes occur at very early stages, and they can be detected via the pattern classification technique applied in this study. (a) All participants; (b) after excluding one clinically normal participant who showed an unusually steep increase in abnormality score over time and subsequently was found to have pathology consistent with Alzheimer's disease despite normal clinical status.

effects regression models were used to compare longitudinal changes in abnormality scores for MCI and cognitively normal groups. Age at assessment and diagnostic status were independent variables and abnormality scores were dependent measures. The MCI group showed a highly significant increase in abnormality score over time ($p < 0.0001$), whereas the cognitively normal group showed a weaker increase over time, $p = 0.022$ (Fig. 3a); the two slopes were significantly different ($p = 0.045$, all one sided tests).

Examination of the individual longitudinal trajectories revealed that one of the clinically normal participants was the main contributor to the longitudinal increase of the abnormality scores of the normal group. We omitted this participant and repeated the mixed effects regression analysis, obtaining the plot of Fig. 3b. Excluding this outlier, there was no significant longitudinal increase of abnormality scores in the normal sample ($p = 0.38$). This participant is now deceased, and autopsy findings indicated moderate plaques and a Braak score of 4, consistent with a pathological diagnosis of possible AD by CERAD criteria (Mirra et al., 1991). This finding is in agreement with his highly abnormal MRI-based classification scores.

Examination of the classification scores on the year of conversion from normal to MCI in the subset of 8 subjects who converted over the course of the study showed that structural brain changes were already quite considerable at that stage. Specifically, the average score of the MCI participants on the year of conversion from normal to MCI was 0.15, whereas the average score of the same participants during their most recent scan during which did not have a diagnosis of dementia was 0.26; the latter scores reflect a more advanced disease stage. The average score of all MCI participants at the latest year of available scans prior to dementia was also 0.26, and the average classification score of all scans of normal controls was −0.3. These numbers indicate that on the year of conversion, the group that progressed to MCI was already well into the range of abnormal brain structure. However, one of the limitations of this study was that we were not able to investigate a variety of specific intervals prior to conversion to MCI due to the modest sample size and the fact that 7 of the MCI individuals were labeled MCI at their first imaging assessment by retrospective ascertainment of date of onset.

3.4. Region of interest (ROI) and voxel-based analysis

We compared the discrimination power of our pattern analysis approach with that of volume measurements of the left and right hippocampus and entorhinal cortex, after normalization by total intracranial volume (ICV) that accounted for total head size. Fig. 4 shows the scatterplot of ICV-normalized volumes, which indicate modest group separation and diagnostic accuracy in these relatively mildly impaired subjects, even though a statistically significant difference between the two groups was measured ($p = 0.036$ for the hippocampus and $p = 0.02$ for the entorhinal cortex, after omission of one outlier; not excluding the outlier yielded respective *p*-
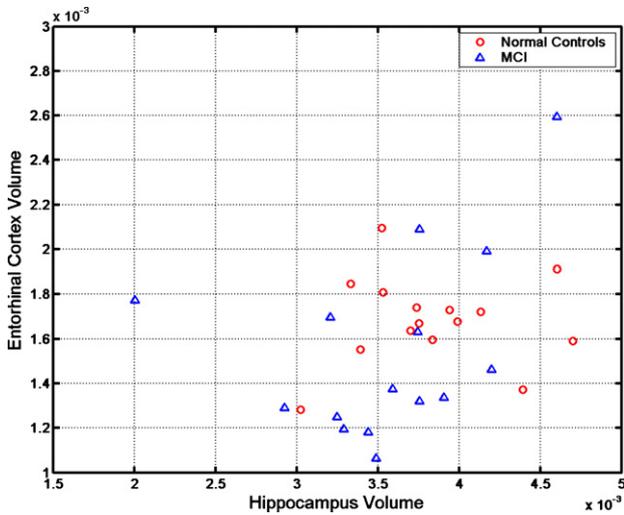
Fig. 4. Scatterplots of the volumes of the hippocampus (horizontal) and entorhinal cortex (vertical), after dividing each measurement by the respective total intra-cranial volume. Although these two structures are known to be significantly affected by Alzheimer's disease, their volumes are of relatively modest diagnostic value, due to the overlap of the two distributions.

Fig. 6. Predictive accuracy as a function of the number of brain regions/clusters included in the classification process. Small number of regions is insufficient, whereas more than 30 regions result in insufficient training of the classifier, as we had 30 samples; this is the well-known "curse of dimensionality" in machine learning, and indicates that more extensive image databases might ultimately improve classification accuracy.

values of 0.08 and 0.14, all one-sided t-tests). In fact, the best classification rate we were able to achieve using the hippocampus and ERC volumes jointly was 76.6%, via a nonlinear support vector machine. The respective ROC is shown in Fig. 5, reflecting a relatively limited diagnostic accuracy.

Comparison of our results with the widely used voxel-based morphometry method, which applied voxel-by-voxel t-tests on the smoothed tissue density maps (Ashburner et al., 2003), revealed that no single brain region carried adequate predictive power. In particular, the classification accuracy of the most discriminatory cluster, examined via cross-validation, was 63.3%, reflecting relatively poor predictive power of single-region measurements, and further

supporting the importance of the network-type of analysis performed herein.

In our final experiment, we investigated the possibility of using a smaller number of brain regions without sacrificing predictive accuracy. Fig. 6 shows the predictive accuracy as a function of the number of clusters, revealing that at least 20 brain regions must be considered jointly, in order to capture the MCI-specific structural pattern.

4. Discussion

To our knowledge, our study is the first to demonstrate that complex and subtle structural patterns that characterize the trajectory of increasing brain abnormalities in individuals with mild MCI can be identified from cross-sectional MR scans via high-dimensional image analysis and pattern classification methods. Importantly, these patterns of structural change can be measured even before cognitive decline brings the individuals to clinical attention. The spatial pattern of distribution of brain tissue that allowed us to accurately detect individuals with the MCI structural phenotype involved several structures that are known to be implicated in AD, including a number of temporal lobe structures, the cingulate, and parts of the orbitofrontal cortex that have dense connections with anterior temporal lobe structures and show pathology early in AD (Bookstein, 2001). Notable is the fact that it was mainly the lateral and anterior part of the hippocampus that contributed to the classification, which might imply that CA1 is relatively more affected in these individuals, and might explain, in part, the relatively limited diagnostic value of total hippocampal volume measurements in the subjects of our study (Fig. 4). A relatively higher CA1 atrophy would also agree with histopathological studies
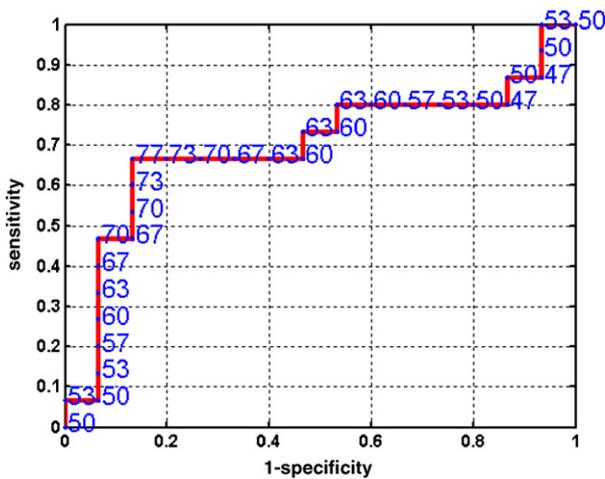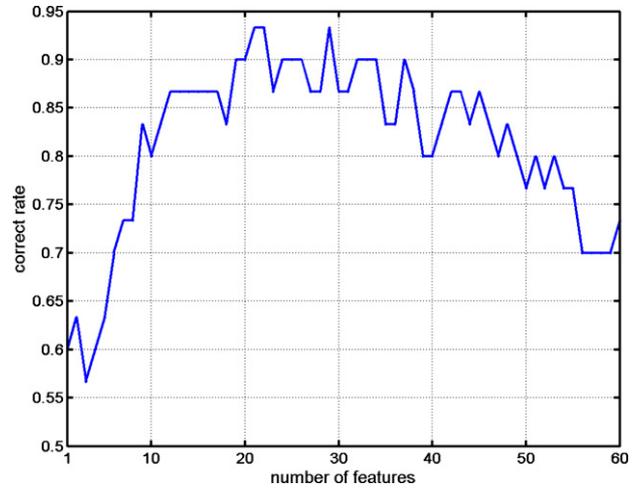

Fig. 5. ROC curve of the classification sensitivity and specificity obtained by using the volumes of the hippocampus and the entorhinal cortex jointly, in conjunction with a nonlinear support vector machine classifier. The resultant diagnostic accuracy is very limited, compared to that of Fig. 2.

(West et al., 2004), albeit the resolution of current 1.5 T MR images does not allow us to specifically define hippocampal subfields. Some recent studies using high-dimensional warping of the hippocampus have also observed spatially heterogeneous patterns of hippocampal atrophy in AD (Frisoni et al., 2006; Wang et al., 2006). These results further bolster our confidence that sophisticated methods for feature selection and pattern classification are necessary to capture the complex patterns of brain atrophy that are most informative for diagnosis, whereas conventional ROI volumetric analysis is limited in that it combines regions that are relatively more affected by disease with regions that are relatively less affected, thereby potentially reducing the predictive power of these measurements. Our analysis also identified white matter regions both in the temporal lobe and in the superior and middle frontal gyri that were very important for accurate classification of MCI and have not been previously reported in the literature. These structures merit further investigation using imaging methods that are more suitable for analysis of white matter structure, especially diffusion tensor imaging. Finally, our results indicated bilateral hippocampal atrophy, in contrast to recent reports of lateralized atrophy, with both greater right hippocampal atrophy (Wang et al., 2006) and greater left hippocampal atrophy (Karas et al., 2004) reported.

Most importantly, our results provide the first evidence that integration of these regional measurements via nonlinear pattern classification provides very high diagnostic accuracy *on an individual basis*. Although several studies examining cross-sectional and longitudinal effects in volumes of brain regions have shown significant group differences between MCI patients that convert to AD and MCI patients that do not convert, or between healthy controls and AD or MCI patients (Bobinski et al., 1999; Convit et al., 1997; Cuenod et al., 1993; De Santi et al., 2001; DeLeon et al., 1991; deToledo-Morrell et al., 1997; Dickerson et al., 2001; Du et al., 2001; Frisoni et al., 1996; Golomb et al., 1993; Jack et al., 1997, 1999; Killiany et al., 1993, 2000a; Krasuski et al., 1998; Laakso et al., 1995, 2000; Lehericy et al., 1994; Rosen et al., 2003; Stoub et al., 2005; Xu et al., 2000), the ability to detect structural patterns that enable accurate prediction for specific individuals is ultimately what determines the clinical value of MRI and measurements obtained from it. Our results further confirmed that evaluation of all these brain regions jointly is necessary to obtain high predictive accuracy. Measurements restricted to volumes of the hippocampus and ERC, two structures that are known to be affected early in AD (Hyman et al., 1984) and show group differences in neuroimaging studies of AD and MCI (Bobinski et al., 1999; Convit et al., 1997; De Santi et al., 2001; deToledo-Morrell et al., 1997; Dickerson et al., 2001; Du et al., 2001; Frisoni et al., 1996; Jack et al., 1999; Killiany et al., 2000a; Krasuski et al., 1998; Lehericy et al., 1994; Rosen et al., 2003; Xu et al., 2000), provided considerably lower predictive accuracy. Lower accuracy was observed even when these measures were evaluated jointly, due to the high overlap of the volume distributions of these two structures between MCI and healthy individuals (Fig. 4).

Analysis of predictive accuracy as a function of the number of brain regions contributing to the classification confirmed that one needs 20–25 clusters to be able to accurately capture all subtleties of the structural abnormality in these mild MCI cases and achieve sufficient predictive accuracy (Fig. 6). Further increase of the number of brain regions reduces predictive accuracy, a fact that is well-known in machine learning and is due to two factors: (1) increase of noise by including regions that are not relevant to classification; (2) insufficient training, as the number of variables exceeds the number of training samples (30).

Clinically important was also the finding that the pattern of spatial distribution of brain tissues characteristic of MCI was typically apparent prior to the recognition of clinical symptoms. In particular, abnormality scores progressed steadily in people that were later diagnosed with MCI, and at the time of conversion to MCI, their scores were well into the abnormal realm. In contrast, individuals who did not develop cognitive impairment showed much slower increase of brain abnormalities during earlier years. Therefore, our findings suggest that there may be a continuum of structural brain changes that eventually reaches some threshold associated with clinically detected cognitive impairment. Furthermore, we have demonstrated that these brain changes can be captured by pattern analysis and summarized by a score reflecting the level of structural abnormality at an individual level. Our results reveal the potential of this approach for early detection of MCI during the prodromal phase of AD, which will be critical in making treatment decisions with the anticipated availability of different treatment options in the future.

Further evaluation of the abnormality scores in the normal group revealed that only one clinically normal participant displayed considerable increase in abnormality score. Autopsy results in this individual demonstrated findings consistent with a pathological diagnosis of AD. Although a single case, this finding bolsters our confidence that our MRI-based classifier is able to detect structural patterns that indicate pathology, even in individuals who are cognitively normal.

The fact that the majority of our MCI participants were only very mildly impaired is both a strength and a possible weakness of our study. Our ability to robustly classify individuals with very mild impairment provides encouragement for future development of this approach as a diagnostic tool for prodromal AD. The majority of our participants showed only mild impairment in memory based on informant ratings. As they were identified within the context of prospective longitudinal follow-ups, they were likely at early stages of impairment and most would not yet have come to clinical attention. On the other hand, we do not know how many of the remaining MCI individuals will ultimately develop AD or when they will do so, and we do not know whether our method will distinguish MCI individuals who remain stable versus those who convert to AD. We also cannot determine at present which cognitively normal individuals with increasing structural abnormality scores will ultimately convert to MCI (Fig. 3) and how quickly they will do so. We continue to fol-

low these participants with cognitive and MRI assessments and have initiated PET amyloid imaging studies in this sample, in order to fully characterize the process of progression from healthy brain structure and cognition to AD.

It is important to emphasize that our sensitivity/specificity estimates were obtained using leave-one-out cross-validation to obtain robust estimates of generalization and predictive power. This ensured that our predictive power was being evaluated on MRI scans not previously encountered by the classifier and the feature selector. This type of cross-validation guards against findings that can potentially be introduced purely by chance, a persistent problem in high-dimensionality analyses especially in relatively small samples. This type of potentially misleading result is illustrated by our observation of 100% separation between MCI and healthy controls in our initial analysis without cross-validation. However, when cross-validated, our analysis showed 90% predictive power, which is an estimate of the diagnostic value of these volumetric measurements on new individuals. While our study demonstrates that individuals with MCI can be distinguished from cognitively healthy controls with high accuracy, sensitivity and specificity, we did not test the potential utility of this approach in differential diagnosis among different types of dementia. A test of the validity of this approach for differential diagnosis of dementia subtypes would require training of the classifier on all types of dementia jointly and merits future investigation.

Our classification results and the estimates of predictive power were derived from a single structural MRI of each individual, rather than from longitudinal measurements of brain atrophy. The ability to accurately classify even mildly impaired individuals from a single cross-sectional MRI contrasts with prevailing thinking that effective prediction of early stages of AD will require measurement of longitudinal brain changes. This is very important from a clinical and financial perspective, since consistent and frequent follow-up of healthy individuals that might or might not be at risk for AD is extremely difficult, especially in a typical clinical setting. We believe that the high-dimensional multivariate nature of our analysis compensated for not using longitudinal measurements for classification. However, if necessary, our classification approach could readily incorporate both cross-sectional and longitudinal measurements, as well as physiological measurements, in order to construct more sensitive and specific early diagnostic tools for AD. It will also be important to continue longitudinal studies to determine how early in the prodromal period these changes are detected.

If our findings are replicated in other samples and imaging centers and, most importantly, if accuracy of classification can be improved by training on more extensive samples compared to the relatively modest number of scans used in this study, our results and the methodology adopted herein can play a significant role in detection of brain abnormalities associated with cognitive impairment prior to recognition of clinical deficits. Moreover, application of our classification methodology to longitudinal imaging data will enrich our understanding of the progression of the underlying brain abnormalities and will be critical in monitoring treatment and guiding therapeutic interventions for one of the most devastating diseases of the elderly.

Disclosure statement

All co-authors have seen and approved this submission. There are no conflicts of interest including any financial, personal or other relationships with other people or organizations, by any of the co-authors, related to the work described in the paper. The submission is not under review by any other archival journal. An oral and poster presentation of this work has been made at the International Conference of Alzheimer's disease, in Madrid, July 2006.

Acknowledgements

References

Ashburner, J., Csernansky, J.G., Davatzikos, C., Fox, N.C., Frisoni, G.B., Thompson, P.M., 2003. Computer-assisted imaging to assess brain structure in healthy and diseased brains. Lancet (Neurology) 2 (2), 79–88.

Bobinski, M., de Leon, M.J., Convit, A., De Santi, S., Wegiel, J., Tarshish, C.Y., Saint Louis, L.A., Wisniewski, H.M., 1999. MRI of entorhinal cortex in mild Alzheimer's disease. Lancet 353 (9146), 38–40.

Bookstein, F.L., 2001. Voxel-based morphometry should not be used with imperfectly registered images. Neuroimage 14, 1454–1462.

Braak, H., Braak, E., Bohl, J., Bratzke, H., 1998. Evolution of Alzheimer's disease related cortical lesions. J. Neural. Transm. 54 (suppl.), 97–106.

Chetelat, G., 2003. Early diagnosis of Alzheiner's disease: contribution of structural neuroimaging. Neuroimage 18 (2), 525–541.

Chetelat, G., Desgranges, B., de la Sayette, V., Viader, F., Eustache, F., Baron, J.-C., 2002. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. Neuroreport 13 (15), 1939–1943.

Convit, A., De Leon, M.J., Tarshish, C., De Santi, S., Tsui, W., Rusinek, H., George, A., 1997. Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. Neurobiol. Aging 18 (2), 131–138.

Convit, A., de Asis, J., de Leon, M.J., Tarshish, C.Y., De Santi, S., Rusinek, H., 2000. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiol. Aging 21 (1), 19–26.

Cuenod, C.A., Denys, A., Michot, J.L., Jehenson, P., Forette, F., Kaplan, D., Syrota, A., Boller, F., 1993. Amygdala atrophy in Alzheimer's disease. An in vivo magnetic resonance imaging study. Arch. Neurol. 50 (9), 941–945.

Davatzikos, C., Acharyya, M., Ruparel, K., Shen, D.G., Loughhead, J., Gur, R.C., Langleben, D., 2005a. Classifying spatial patterns of brain activity for lie-detection. Neuroimage 28 (3), 663–668.

Davatzikos, C., Shen, D.G., Wu, X., Lao, Z., Hughett, P., Turetsky, B.I., Gur, R.C., Gur, R.E., 2005b. Whole-brain morphometric study of schizophrenia reveals a spatially complex set of focal abnormalities. JAMA Arch. Gen. Psychiatry 62, 1218–1227.

De Santi, S., de Leon, M.J., Rusinek, H., Convit, A., Tarshish, C.Y., Roche, A., Tsui, W.H., Kandil, E., Boppana, M., Daisley, K., Wang, G.J., Schlyer, D., Fowler, J., 2001. Hippocampal formation glucose metabolism and volume losses in MCI and AD. Neurobiol. Aging 22 (4), 529–539.

DeLeon, M.J., George, A.E., Golomb, J., Tarshish, C., Convit, A., Kluger, A., DeSanti, S., Mcrae, T., Ferris, S.H., Reisberg, B., Ince, C., Rusinek, H., Bobinski, M., Quinn, B., Miller, D.C., Wisniewski, H.M., 1991. Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. Neurobiol. Aging 18, 1–11.

deToledo-Morrell, L., Sullivan, M.P., Morrell, F., Wilson, R.S., Bennett, D.A., Spencer, S., 1997. Alzheimer's disease: in vivo detection of differential vulnerability of brain regions. Neurobiol. Aging 18, 463–538.

Dickerson, B.C., Goncharova, I., Sullivan, M.P., Forchetti, C., Wilson, R.S., Bennett, D.A., Beckett, L.A., deToledo-Morrell, L., 2001. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiol. Aging 22 (5), 747–754.

Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., Yaffe, K., Kramer, J.H., Reed, B., Norman, D., Chui, H.C., Weiner, M.W., 2001. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 71 (4), 441–447.

Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. In: Duncan, J.S., Gerig, G. (Eds.), Lecture Notes in Computer Science,. Springer-Verlag GmbH, Palm Springs, CA, USA, pp. 1–8.

Frisoni, G.B., Beltramello, A., Weiss, C., Geroldi, C., Bianchetti, A., Trabucchi, M., 1996. Linear measures of atrophy in mild Alzheimer disease. AJNR Am. J. Neuroradiol. 17 (5), 913–923.

Frisoni, G.B., Sabattoli, F., Lee, A.D., Dutton, R.A., Toga, A.W., Thompson, P.M., 2006. In vivo neuropathology of the hippocampal formation in AD: a radial mapping MR-based study. NeuroImage 32 (1), 104–110.

Goldszal, A.F., Davatzikos, C., Pham, D., Yan, M., Bryan, R.N., Resnick, S.M., 1998. An image processing protocol for the analysis of MR images from an elderly population. J. Comput. Assist. Tomogr. 22 (5), 827–837.

Golland, P., Fischl, B., Spiridon, M., Kanwisher, N., Buckner, R.L., Shenton, M.E., Kikinis, R., Dale, A., Grimson, W.E.L., 2002. In: Dohi, T., Kikinis, R. (Eds.), Discriminative Analysis for Image-based Studies. Springer-Verlag GmbH, Tokyo, Japan, pp. 508–515.

Golomb, J., deLeon, M.J., Kluger, A., George, A.E., Tarshish, C., Ferris, S.H., 1993. Hippocampal atrophy in normal aging: an association with recent memory impairment. Arch. Neurol. 50 (9), 967–973.

Grundman, M., et al., 2004. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. Arch. Neurol. 61 (1), 59–66.

Hyman, B.T., Van Horsen, G.W., Damasio, A.R., Barnes, C.L., 1984. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. Science 225 (4667), 1168–1170.

Jack Jr., C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 49 (3), 786–794.

Jack, C.R., Petersen, R.C., Xu, Y.C., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E., Kokmen, E., 1999. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology 52 (7), 1397–1403.

Karas, G.B., Scheltens, P., Rombouts, S.A.R.B., Visser, P.J., Schijndel RAv, Fox, N.C., Barkhof, F., 2004. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. Neuroimage 23 (2), 708–716.

Kaye, J., Swihart, T., Howieson, D., Dame, A., Moore, M., Karnos, T., Camicioli, R., Ball, M., Oken, B., Sexton, G., 1997. Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. Neurology 48 (5), 1297–1304.

Killiany, R.J., Moss, M.B., Albert, M.S., Sandor, T., Tieman, J., Jolesz, F., 1993. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch. Neurol. 50 (9), 949–954.

Killiany, R.J., Gomez-Isla, T., et al., 2000a. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. Ann. Neurol. 47 (4), 430–439.

Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., Albert, M.S., 2000b. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Ann. Neurol. 47 (4), 430–439.

Krasuski, J.S., Alexander, G.E., Horwitz, B., Daly, E.M., Murphy, D.G., Rapoport, S.I., Schapiro, M.B., 1998. Volumes of medial temporal lobe structures in patients with Alzheimer's disease and mild cognitive impairment (and in healthy controls). Biol. Psychiatry 43 (1), 60–68.

Laakso, M.P., Soininen, H., Partanen, K., Helkala, E.L., Hartikainen, P., Vainio, P., Hallikainen, M., Hanninen, T., Riekkinen Sr., P.J., 1995. Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions. J. Neural. Transm. Park. Dis. Dement. Sect. 9 (1), 73–86.

Laakso, M.P., Hallikainen, M., Hanninen, T., Partanen, K., Soininen, H., 2000. Diagnosis of Alzheimer's disease: MRI of the hippocampus versus delayed recall. Neuropsychologia 38 (5), 579–584.

Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21 (1), 46–57.

Lehericy, S., Baulac, M., Chiras, J., Pierot, L., Martin, N., Pillon, B., Deweer, B., Dubois, B., Marsault, C., 1994. Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. AJNR Am. J. Neuroradiol. 15 (5), 929–937.

Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J., Lopez, O., Meltzer, C., 2004. In: Barillot, C., Haynor, D.R., Hellier, P. (Eds.), Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Feature Classification. Springer-Verlag GmbH, Saint-Malo, France, pp. 393–401.

Medina, D., deToledo-Morrell, L., Urresta, F., Gabrieli, J.D.E., Moseley, M., Fleischman, D., Bennett, D.A., Leurgans, S., Turner, D.A., Stebbins, G.T., 2006. White matter changes in mild cognitive impairment and AD: a diffusion tensor imaging study. Neurobiol. Aging 27 (5), 663–672.

Mirra, S., Heyman, A., McKeel, D., Sumi, S., Crain, B., Brownlee, L., Vogel, F., Hughes, J., van Belle, G., Berg, L., 1991. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41 (4), 479–486.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., Clark, C., 1989. The consortium to establish a registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 39 (9), 1159–1165.

Nestor, P.J., Scheltens, P., Hodges, J.R., 2004. Advances in the early detection of Alzheimer's disease. Nat. Rev. Neurosci. 5, S34–S41.

Petersen, R.C., 2003. Mild cognitive impairment: aging to Alzheimer's Disease. Oxford University Press.

Resnick, S.M., Goldszal, A., Davatzikos, C., Golski, S., Kraut, M.A., Metter, E.J., Bryan, R.N., Zonderman, A.B., 2000. One-year age changes in MRI brain volumes in older adults. Cereb. Cortex 10, 464–472.

Rosen, A.C., Prull, M.W., Gabrieli, J.D., Stoub, T., O'Hara, R., Friedman, L., Yesavage, J.A., deToledo-Morrell, L., 2003. Differential associations between entorhinal and hippocampal volumes and memory performance in older adults. Behav. Neurosci. 117 (6), 1150–1160.

Shen, D.G., Davatzikos, C., 2003. Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. Neuroimage 18 (1), 28–41.

Stoub, T.R., Bulgakova, M., Leurgans, S., Bennett, D.A., Fleischman, D., Turner, D.A., deToledo-Morrell, L., 2005. MRI predictors of risk of incident Alzheimer disease: a longitudinal study. Neurology 64 (9), 1520–1524.

Vapnik, V.N., 1998. Statistical Learning Theory. Wiley, New York, 736 pp.

Wang, L., Miller, J.P., Gado, M.H., McKeel, D.W., Rothermich, M., Miller, M.I., Morris, J.C., Csernansky, J.G., 2006. Abnormalities of hippocampal surface structure in very mild dementia of the Alzheimer type. NeuroImage 30 (1), 52–60.

West, M., Kawas, C., Stewart, W., Rudow, G., Troncoso, J., 2004. Hippocampal neurons in pre-clinical Alzheimer's disease. Neurobiol. Aging 25 (9), 1205–1212.

Xu, Y., Jack Jr., C.R., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. Neurology 54 (9), 1760–1767.