



Joint Correlational and Discriminative Ensemble Classifier Learning for Dementia Stratification Using Shallow Brain Multiplexes

Rory Raeper¹, Anna Lisowska², Islem Rezik¹(✉),
and The Alzheimer's Disease Neuroimaging Initiative

¹ BASIRA lab, CVIP group, School of Science and Engineering, Computing,
University of Dundee, Dundee, UK

irekik@dundee.ac.uk

² Department of Computer Science, The University of Warwick, Coventry, UK
<http://www.basira-lab.com>

Abstract. The demented brain wiring undergoes several changes with dementia progression. However, in early dementia stages, particularly early mild cognitive impairment (eMCI), these remain challenging to spot. Hence, developing accurate diagnostic techniques for eMCI identification is critical for early intervention to prevent the onset of Alzheimer's Disease (AD). There is a large body of machine-learning based research developed for classifying different brain states (e.g., AD vs MCI). These works can be fundamentally grouped into two categories. The first uses *correlational* methods, such as canonical correlation analysis (CCA) and its variants, with the aim to identify most correlated features for diagnosis. The second includes *discriminative* methods, such as feature selection methods and linear discriminative analysis (LDA) and its variants to identify brain features that distinguish between two brain states. However, existing methods examine these correlational and discriminative brain data *independently*, which overlooks the complementary information provided by both techniques, which could prove to be useful in the classification of patients with dementia. On the other hand, how early dementia affects cortical brain *connections in morphology* remains largely unexplored. To address these limitations, we propose a joint correlational and discriminative ensemble learning framework for eMCI diagnosis that leverages a novel brain network representation, derived from the cortex. Specifically, we devise 'the shallow convolutional brain multiplex' (SCBM), which not only measures the similarity in morphology between pairs of brain regions, but also encodes the relationship between two morphological brain networks. Then, we represent each individual brain using a set of SCBMs, which are used to train joint ensemble CCA-SVM and LDA-based classifier. Our framework outperformed several state-of-the-art methods by 3-7% including independent correlational and discriminative methods.

1 Introduction

Early mild cognitive impairment (eMCI) is an early stage of dementia, that affects brain function and cognition in subtle ways that remain challenging to spot when mapping brain connections using Magnetic Resonance Imaging (MRI) in the disordered brain. Undoubtedly, understanding how early dementia alters specific brain connections across different patients might help better diagnose and stratify early stages of brain dementia, treat patients effectively, and eventually slow down worsening of symptoms and conversion to Alzheimer’s Disease (AD). Within this scope, several machine learning approaches leveraged multi-modal (MRI) data including resting-state functional MRI (rsfMRI) and diffusion MRI (dMRI) to distinguish between patients with MCI and healthy controls [1]. However, the very early brain states of dementia including eMCI remain least investigated in dementia literature, compared with AD and MCI states.

Recent machine-learning methods were devised for MCI identification using connectomic brain data [2,3]. However, existing works mainly used functional networks (derived from rsfMRI) and structural networks (derived from dMRI). These exclude the recent landmark works [4–6], which devised morphological brain networks (MBN) for mapping morphological ‘connections’ in the cortex. Basically, an MBN is generated by measuring the difference in morphology between two cortical regions based on a specific cortical attribute (e.g., sulcal depth). More importantly, [4,6] proposed to embed multiple brain networks into a multiplex network structure composed of intra-layer and inter-layer networks. Each intra-layer network in the multiplex represents an MBN derived from a specific cortical attribute, whereas an inter-layer network is a network-to-network similarity slid between two consecutive intra-layers. The integrated inter-layer network is able to capture high-order brain alterations at the morphological level. While [6] used correlational inter-layers in the brain multiplex structure for late dementia diagnosis, [4] proposed convolutional inter-layers produced by convolving two consecutive MBNs (intra-layers) in the multiplex for early dementia stratification. Notably, both multiplex architectures outperformed conventional single-layer and multi-layer brain network representations. Furthermore, while [6] used a machine learning method that identifies discriminative connective features for dementia classification, [4] proposed a correlation-based ensemble learning framework, which identifies highly correlated multiplex features. Such approaches disentangle correlational from discriminative approaches, which might limit our understanding of disordered connective changes in the diseased brain.

Broadly, existing classification approaches can be categorized into two groups: (1) methods that aim to identify highly correlated features such as Canonical Correlation Analysis (CCA) [4,7,8], and (2) methods that seek to identify the most discriminative features using feature selection methods such as [9] or discriminative analysis [10]. The first group includes all related CCA works and their variants such as sparse CCA (sCCA) [11] and non-linear kernel CCA (kCCA) [12]. Typically, CCA maps input features into a shared space where their correlation is maximized, and the mapped features can then be fused. The second

group comprises discriminative machine learning approaches, such as Linear Discriminant Analysis (LDA), where the input features are projected onto a space where their disparity and discriminability are maximized [10]. Other methods integrate a discriminative feature selection method such as mutual information (MutInf-FS) [9] and Infinite Feature Selection (Inf-FS) [6, 13]. However, a fundamental limitation of the above methods and works reviewed in [14] consists in either identifying correlational features or discriminative features for stratifying dementia states. This overlooks the complementary information that can be integrated from both correlational and discriminative approaches to further the eMCI/NC classification accuracy.

To fill this gap, we propose a joint correlational and discriminative ensemble learning framework, which first pairs multi-source brain multiplex data generated from a set of MBNs. Next, each pair is communicated to two different blocks of our framework: the first block including a set of discriminative classifiers and the second block including a set of correlational classifiers. Ultimately, we aggregate labels predicted by both blocks using majority voting to output the final label for a target testing subject. In addition to this landmark contribution, we propose a novel multi-layer brain network architecture, the shallow convolutional brain multiplex (SCBM), which unlike the deep CBM proposed in [4], is generated using only two MBNs. This avoids creating redundant features when pairing multiplexes prior to passing them forward to classifiers.

2 Ensemble LDA and CCA-SVM Paired Classifier Learning using Shallow Convolutional Brain Multiplexes

In this section, we introduce the concept of a shallow convolutional brain multiplex and present our novel joint correlational and discriminative ensemble learning framework. Fig. 1 shows the different steps for (A) shallow convolutional brain multiplex construction from cortical surface, and (B) multi-source SCBM data pairing for training the correlational block comprising a set of CCA-based SVM classifiers and the discriminative block including a set of LDA classifiers. Below we detail the different steps of our eMCI/NC classification framework.

Single-View Morphological Brain Network (MBN) Construction. For each cortical attribute (e.g., cortical thickness), we construct a single-view network for each subject. Such network comprises a set of nodes (anatomical brain regions) and a collection of edges interconnecting the nodes (representing the difference between the two brain regions in *morphology*). The average value of a cortical attribute was calculated for each anatomical region of interest (ROI). For each cortical attribute, the strength of each network edge connecting two ROIs is then computed as the absolute difference between their average values, thereby quantifying their dissimilarity (Fig. 1). The same procedure was followed to obtain the connectivity matrices from different cortical attributes (e.g., sulcal depth, curvature) [4, 6].

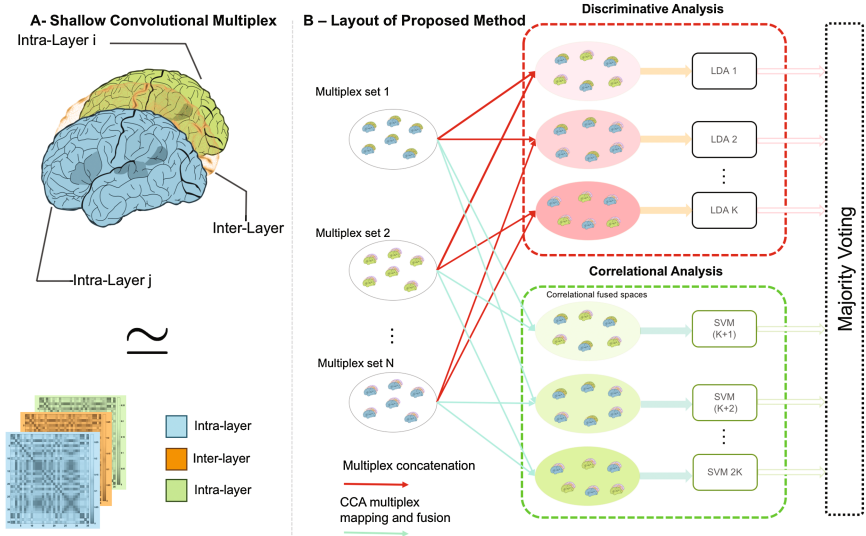


Fig. 1. Pipeline of the proposed joint correlational and discriminative ensemble learning using shallow convolutional brain multiplexes. (A) shows the construction of a single multiplex where the inter-layers are created between two intra-layers (two MBNs derived from the cortical surface). (B) We first represent each subject using N multiplexes, produced using different combinations of morphological brain networks. Next, for all possible combinations of multiplex pairs, each pair of multiplexes is passed into the ensemble framework, consisting of a correlational learning block (where they are mapped by CCA and classified by SVM) and a discriminative block (where they are mapped and separated into two classes by LDA). The two blocks produce predicted class labels for the test subjects based on analysis of subsequent pairs of multiplexes. The final class label is assigned through majority voting on labels assigned by the two blocks.

Convolutional Brain Multiplex Construction. In a generic way, we define a brain multiplex \mathcal{M} using a set of M intra-layers (or MBNs) $\{\mathbf{V}_1, \dots, \mathbf{V}_M\}$, each representing a single view of the brain morphology (i.e., cortical attribute), where between two consecutive intra-layers \mathbf{V}_i and \mathbf{V}_j we slide an inter-layer $\mathbf{C}_{i,j}$, which is defined by convolving two consecutive intra-layers. Convolution captures the signal within a subgraph (a small patch in the connectivity matrix) extracted from a first layer (whole matrix) as an expression of other subgraphs extracted from a second layer. One can think of the inter-layer network as a ‘blending’ of both intra-layers, expressing the amount of overlap of first intra-layer as it is shifted over the second intra-layer.

Each element in row a and column b within the convolutional inter-layer matrix $\mathbf{C}_{i,j}$ between views \mathbf{V}_i and \mathbf{V}_j is defined as: $\mathbf{C}_{i,j}(a,b) = \sum_p \sum_q \mathbf{V}_i(p,q) \mathbf{V}_j(a-p+1, b-q+1)$. The multiplex architecture allows not only to explore how different brain views get altered by a specific disorder, but how their relationship might get affected. Since the morphological brain

connectivity matrices are symmetric (**Fig. 1–A**), we extract features from each MBN by directly concatenating the off-diagonal weights of all connectivities in each upper triangular matrix. For each network of size $n \times n$, we extract a feature vector of size $(n \times (n - 1)/2)$. Previously, in [4], the generalized multiplex architecture was proposed: $\mathcal{M} = \{\mathbf{V}_1, \mathbf{C}_{1,2}, \mathbf{V}_2, \dots, \mathbf{V}_j, \mathbf{C}_{i,j}, \mathbf{V}_j, \dots, \mathbf{V}_M\}$. Next, to capture the inter-relationship between all possible combinations of intra-layers in a multiplex, a set of N multiplexes were generated for each subject through reordering the intra-layer networks, thereby generating an *ensemble of brain multiplexes* $\mathbb{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_N\}$. However, this approach resulted in many highly correlated features used for the ensemble learning, which may somewhat mislead classifier learning. To minimize the correlation between different multiplexes when pairing them for ensemble classifier training, we propose a shallow (i.e., 2-layer) convolutional brain multiplex structure. We define a *shallow* multiplex $\mathcal{M} = \{\mathbf{V}_i, \mathbf{C}_{i,j}, \mathbf{V}_j\}$ using 2 intra-layers \mathbf{V}_i and \mathbf{V}_j and an inter-layer $\mathbf{C}_{i,j}$ encoding the relationship between \mathbf{V}_i and \mathbf{V}_j , slid in between them (**Fig. 1–A**). We note that each subject-specific brain multiplex \mathcal{M} in \mathbb{M} captures unique similarities between two different morphological brain network views (e.g., sulcal depth network and cortical thickness network) that are not present in a different shallow multiplex.

Proposed Joint Canonical Correlational and Discriminative Mappings of SCBM Sets. Since each multiplex $\mathcal{M}_k \in \mathbb{M}$ captures a unique and complex relationship between different brain network views, one needs to examine all morphological brain multiplexes in the ensemble \mathbb{M} . This will provide us with a more holistic understanding of how explicit morphological brain connections can be altered by dementia onset as well as how their implicit high-order (a connection of connections) relationship can be affected. To make use of all the information available from different multiplexes, in the *correlational learning block* of our framework (outlined in green **Fig. 1–B**), we use CCA [7,8] to map pairs of multiplex features extracted from different sets into a shared subspace that depicts highly-correlated relevant features. We then concatenate the CCA-mapped multiplex features from the first and second sets. This correlational block allows to minimize the multiplex set-specific noise and reduces multiplex data dimensionality. Next, we use each CCA-mapped pair of multiplex features $\tilde{\mathbf{M}}_{k,l}^c$ to train a linear support vector machine (SVM) classifier (**Fig. 1–B**). Noting that for each training subject we have N multiplexes estimated, we perform C_N^2 mappings of each SCBM pair in \mathbb{M} .

Simultaneously, we train the *paralleled discriminative block* (outlined in red **Fig. 1–B**) aggregating sets of regularized LDA classifiers using the paired SCBM features from different sets in a *supervised* manner. Specifically, each LDA classifier attempts to maximize the difference between multiplex features so that there are distinct groups based on the given class labels. All training multiplex features are mapped into a discriminative space guided by the labels, where discriminative paired multiplex features are generated $\tilde{\mathbf{M}}_{k,l}^d$. In the testing stage, we use the learned correlational and discriminative transformations to respectively map each pair of testing multiplex feature vector onto their corresponding

CCA space where they are communicated to an SVM classifier and LDA space, respectively. Finally, to identify the label of the testing subject, we use majority voting by selecting the highly frequent predicted label outputted by classifiers in both blocks. We note that LDA performs both feature dimensionality reduction and classification, while CCA only maps the features, thus requires to be combined with a classifier such as SVM.

3 Results and Discussion

Data. We used leave-one-out cross validation to evaluate the proposed classification framework on 82 subjects (42 eMCI and 42 NC) from ADNI GO public dataset¹, each with structural T1-w MR image. We used FreeSurfer [15] to reconstruct both right and left cortical surfaces for each subject from T1-w MRI. Then we parcellated each cortical hemisphere into 35 cortical regions using Desikan-Killiany Atlas. For the deep CBM, we defined $N = 6$ multiplexes, each using $M = 4$ MBNs, anchored at \mathbf{V}_1 . For each cortical attribute (signal on the cortical surface), we compute the strength of the morphological network connection linking i^{th} ROI to the j^{th} ROI as the absolute difference between the averaged attribute values in both ROIs. Multiplex \mathcal{M}_1 includes cortical attribute views $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4\}$, \mathcal{M}_2 includes $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_4, \mathbf{V}_3\}$, \mathcal{M}_3 includes $\{\mathbf{V}_1, \mathbf{V}_3, \mathbf{V}_4, \mathbf{V}_2\}$, \mathcal{M}_4 includes $\{\mathbf{V}_1, \mathbf{V}_3, \mathbf{V}_2, \mathbf{V}_4\}$, \mathcal{M}_5 includes $\{\mathbf{V}_1, \mathbf{V}_4, \mathbf{V}_2, \mathbf{V}_3\}$, and \mathcal{M}_6 includes $\{\mathbf{V}_1, \mathbf{V}_4, \mathbf{V}_3, \mathbf{V}_2\}$. For each cortical region, \mathbf{V}_1 denotes the maximum principal curvature brain view, \mathbf{V}_2 denotes the mean cortical thickness brain view, \mathbf{V}_3 denotes the mean sulcal depth brain view, and \mathbf{V}_4 denotes the mean average curvature brain view. As for the proposed SCBM, we define $N = C_4^2 = 6$ shallow multiplexes by considering all possible pairings of 2 views out of 4. For our experiments, we created 4 representations of MBN data: (1) ‘Views’ by concatenating all MBNs, (2) ‘Correlational multiplexes’ with inter-layer computed using Pearson correlation, (3) ‘Convolutional multiplexes’ composed of 4 intra-layers with inter-layers generated using 2D convolution, and (4) ‘Shallow convolutional multiplexes’ composed of 2 intra-layers with inter-layers generated using 2D convolution.

Comparison Methods and Evaluation. To demonstrate the effectiveness of integrating correlational and discriminative methods into a single framework, we benchmarked our method against several discriminative methods including: Eigenvector Centrality (ECFS) [16], Mutual Information (MutInf-FS) [17], and Infinite Feature Selection (Inf-FS) [13]. We also benchmarked our method against the CCA-based eMCI/NC classification framework in [4]. We also evaluated the performance of each of the aforementioned discriminative methods when combined with CCA using our proposed framework using MBNs derived from the right hemisphere since significantly greater cortical atrophy is observed in the right hemisphere of MCI patients compared with the left hemisphere [18]. A leave-one-out (LOO) cross-validation (CV) scheme was used to test all these

¹ <http://adni.loni.usc.edu>.

Table 1. Average eMCI/NC classification accuracy using our method and different comparison methods.

Method	Dataset	Accuracy	Sensitivity(%)	Specificity(%)
Ensemble SVM Paired Classifiers using CCA[4]	Views	67.86	61.9	73.81
	Correlational	58.33	64.29	52.38
	Convolutional	71.43	73.81	69.05
	Shallow Conv	73.81	76.19	71.43
Ensemble SVM Paired Classifiers using ECFS[14]	Views	73.81	64.29	83.33
	Correlational	73.81	73.81	73.81
	Convolutional	76.19	73.81	78.57
	Shallow Conv	66.67	66.67	66.67
Ensemble SVM Paired Classifiers using CCA[6]+ ECFS[14]	Views	69.05	64.29	73.81
	Correlational	57.14	61.9	52.38
	Convolutional	70.24	73.81	66.67
	Shallow Conv	78.57	78.57	78.57
Ensemble SVM Paired Classifiers using MutInf-FS[15]	Views	72.62	66.67	78.57
	Correlational	63.1	61.9	64.29
	Convolutional	64.29	64.29	64.29
	Shallow Conv	76.19	78.57	73.81
Ensemble SVM Paired Classifiers Using CCA[6]+ MutInf-FS[15]	Views	66.67	61.9	71.43
	Correlational	54.76	57.14	52.38
	Convolutional	71.43	71.43	71.43
	Shallow Conv	77.38	78.57	76.19
Ensemble LDA Paired Classifiers[9]	Views	70.24	61.9	78.57
	Correlational	71.43	71.43	71.43
	Convolutional	77.38	78.57	76.19
	Shallow Conv	73.81	73.81	73.81
Ensemble LDA[9] and CCA-SVM[6] Paired Classifiers (Ours)	Views	70.24	69.05	71.43
	Correlational	70.24	66.67	73.81
	Convolutional	79.76	78.57	80.95
	Shallow Conv	80.95	83.33	78.57

methods, with a 5-fold nested CV to optimize the number of selected features for discriminative methods. Furthermore, each of these methods was evaluated using the 4 representations of MBN data.

Best performance. Table 1 displays the results for our proposed framework and all comparison methods. Overall, merging discriminative and correlational methods in an ensemble learning framework consistently outperformed the base methods when used independently. Furthermore, our method, combining CCA and LDA, achieved the best classification accuracy 80.95% using shallow convolutional

brain multiplexes. Compared with other correlational-discriminative frameworks (e.g., CCA + ECFS) and the recent work [4], our method increased the classification accuracy by $\sim 3\text{-}7\%$.

Shallow vs. deep convolutional brain multiplexes. The proposed SCBM consistently outperformed concatenated MBN views and correlation brain multiplexes across all methods –except for independent ECFS. Since different deep multiplexes contain overlapping sets of features, resulting in highly-correlated input data, it might result in a suboptimal ensemble performance. Hence, the new shallow multiplex structure solved this problem by reducing the correlation between individual classifiers in the ensemble and overall produced a better ensemble classifier performance compared to the ensemble classifier using deep convolutional multiplex structure [4].

4 Conclusion

Diagnosing early brain symptoms of dementia such as early Mild Cognitive Impairment (eMCI) is vital to prevent worsening of symptoms. To assist this diagnosis, we proposed a joint correlational and discriminative ensemble learning framework using shallow convolution brain multiplexes. Our method attained a large increase in accuracy when using both the shallow and deep convolutional data against several benchmark methods including [4], and numerous discriminative methods. A reported increase of over 7% was attained for the shallow data which supports our theory that utilizing both correlational and discriminative analysis methods yields an increase in overall performance. Another conclusion drawn from these results is the similar accuracy between the shallow and deep convolutional data with the shallow having a higher prediction accuracy frequently. This shows that investigating the similarity between two brain networks can be convenient when analyzing the multi-level effects dementia has on brain connections. Future work may integrate genomic, functional and structural networks as well as explore a wider variety of discriminative feature selection methods together with a broad array of correlational methods (such as Sparse CCA or Kernel CCA) to explore.

References

1. Brown, C.J., Hamarneh, G.: Machine learning on human connectome data from MRI. arXiv preprint [arXiv:1611.08699](https://arxiv.org/abs/1611.08699) (2016)
2. Chen, X., Zhang, H., Gao, Y., Wee, C.Y., Li, G., Shen, D.: High-order resting-state functional connectivity network for MCI classification. *Human brain mapping* **37**, 3282–3296 (2016)
3. Wee, C.Y., Yang, S., Yap, P.T., Shen, D., Initiative, A.D.N., et al.: Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain imaging and behavior* **10**, 342–356 (2016)
4. Lisowska, A., Rekić, I., Initiative, A.D.N., et al.: Pairing-based ensemble classifier learning using convolutional brain multiplexes and multi-view brain networks for early dementia diagnosis. *Springer* (2017) 42–50

5. Soussia, M., Rekik, I.: High-order connectomic manifold learning for autistic brain state identification. *International Workshop on Connectomics in Neuroimaging* (2017) 51–59
6. Mahjoub, I., Mahjoub, M.A., Rekik, I.: Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states. *Scientific reports* **8**, 4103 (2018)
7. Zhu, X., Suk, H.I., Lee, S.W., Shen, D.: Canonical feature selection for joint regression and multi-class identification in alzheimer’s disease diagnosis. *Brain imaging and behavior* **10**, 818–828 (2016)
8. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: An overview with application to learning methods. *Neural computation* **16**, 2639–2664 (2004)
9. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence* **27**, 1226–1238 (2005)
10. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.* (1999) 41–48
11. Hardoon, D.R., Shawe-Taylor, J.: Sparse canonical correlation analysis. *Machine Learning* **83**, 331–353 (2011)
12. Akaho, S.: A kernel method for canonical correlation analysis. *arXiv preprint cs/0609071* (2006)
13. Roffo, G., Melzi, S., Cristani, M.: Infinite feature selection. *Proceedings of the IEEE International Conference on Computer Vision* (2015) 4202–4210
14. Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., Davatzikos, C.: A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages. *NeuroImage* **155**, 530–548 (2017)
15. Fischl, B.: Freesurfer. *Neuroimage* **62**, 774–781 (2012)
16. Roffo, G., Melzi, S.: Features selection via eigenvector centrality. *Proceedings of New Frontiers in Mining Complex Patterns (NFMCP 2016)(Oct 2016)* (2016)
17. Bennasar, M., Hicks, Y., Setchi, R.: Feature selection using joint mutual information maximisation. *Expert Systems with Applications* **42**, 8520–8532 (2015)
18. Apostolova, L.G., Steiner, C.A., Akopyan, G.G., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M.: Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. *Archives of neurology* **64**, 1489–1495 (2007)