A support vector machine-based method to identify subjective cognitive decline with multi-modal imaging

Introduction

Subjective cognitive decline (SCD), a decline in cognition that lacks so-called 'objective evidence'[1] in cognitive tests, characterized by increasing compensatory cognitive efforts and subtle cognitive decline[2], is identified by recent studies as a preclinical stage of Alzheimer's disease (AD), a progressive, irreversible neurodegenerative disease characterized by memory impairment, decline in abstract thinking and computation, change in personality and corresponding actions, and other cognitive dysfunctions. Indicated by researches[3,4,5,6,7] as a risk factor for mild cognitive impairment (MCI), a cognitive impairment generally regarded as a transitional period between normal aging and AD[8], SCD serves as a symptomatic indicator of preclinical AD. Because early initiation before the irreversible brain tissue impairment caused by AD is essential for efficient AD treatments[9], it is very important to find an effective method to distinguish SCD patients from normal controls (NCs), which may have useful application in AD diagnoses and treatments.

Machine learning as a rising field was recently shown to possess the capability to be applied to automatic classification of different groups of subjects with distinct features. Multi-modal imaging, followed by machine learning methods, was found to have the potentiality of differentiating MCI patients from healthy subjects. In order to achieve this goal, classification algorithms were applied to diffusion tensor imaging (DTI) and positron emission tomography (PET)[10,11,12]. Structural magnetic resonance measurement of the atrophy of gray matter in MCI and AD found that the first brain region to be affected by atrophy is the medial temporal lobe containing the hippocampus (HIP), parahippocampal gyrus (PHG) and amygdala (AMYG)[13,14]. Some studies have employed sMRI characteristics, such as voxel-wise volume or vertex-based cortical thickness features, to identify MCI/AD patients from NCs[15,16,17]. Resting-state functional MRI (fMRI) provides a primary method of mechanism detection, diagnostic assessment or therapeutic monitoring of MCI and AD[18,19]. Among the studies that applied classification methods to fMRI data, Zhang et al. adopted the regional homogeneity (ReHo) characteristic of fMRI as the classification features, and achieved an accuracy of 71.4%[20], Chen et al. used Fisher linear discriminant analysis based on the coefficients of the large-scale network, and yielded an accuracy of 91%[21]. Several recent studies have reported that making information integrated effectively from multi-modal imaging can significantly improve the classification performance[22,23]. Wee et al. and Dyrba et al. used a multi-kernel support vector machine (SVM) for multi-modal imaging classification of MCI and AD subjects, and both of them obtained high accuracies[24]. However, few studies tried to apply machine learning methods to identify SCD subjects from NCs. Considering the promising results from utilizing machine learning into differentiating MCI patients from NCs, we intend to further apply this method into the classification of SCD subjects and NCs.

In this paper, we proposed a SVM-based machine learning classification model to discriminate SCD subjects from NCs. The model adopted classification features from multi-modal imaging of diffusion tensor imaging (DTI) resting-state functional magnetic resonance imaging (rs-fMRI), applied four different method, including T-test, group lasso, elastic and dirty model, to acquire an effective algorithm.

Method

To extract data from two modalities (DTI and rs-fMRI), we utilized brain network features; conducted feature selections by sparse method, including Lasso, dirty model, and extended dirty model; and implemented a multimodality data fusion using multiple-kernel SVM by methods of linear, rbf, and poly.

1. Subject & data acquisition

A total of 224 right-hand, Han Chinese subjects were enrolled in this study from September 2009 to December 2015. Among them, 62 NC subjects were recruited from the local community by advertisements. 162 subjects with memory concerns were recruited from the memory clinic of the Neurology Department of Xuan Wu Hospital in Beijing, China, including 47 SCD, 60 aMCI, and 55 d-AD subjects. The study was approved by the Medical Research Ethics Committee and Institutional Review Board of Xuan Wu Hospital (Clinical Trials. gov Identifier: NCT02353884 and NCT02225964). All subjects were provided with written informed consent and signed it prior to any experimental procedures. All subjects underwent a series of standardized clinical evaluations, including a medical history interview, neurologic examination, and a string of neuropsychological tests. Neuropsychological tests included the Chinese version of the Mini–Mental State Examination (MMSE), the Beijing version of Montreal Cognitive Assessment (MoCA)[25], Clinical Dementia Rating Scale (CDR) 31, the auditory verbal learning test (AVLT)[26], Activity of Daily Living (ADL), Hachinski ischemic scale, Hamilton depression rating scale (HAMD) [27], and the Center for Epidemiologic Studies depression scale et al.[28]. The diagnosis was performed by experienced neurologists. All the subjects were diagnosed according to guidelines and asked to fulfill a brain Magnetic Resonance Imaging (MRI) scan once they were enrolled.

The NC patients must meet the research criteria: (a) without memory concerns; (b) MMSE and MoCA scores were within the normal range (adjusted for age, sex, and education); and (c) CDR score of 0.

The diagnosis of SCD fulfilled published SCD research criteria proposed by Subjective Cognitive Decline Initiative (SCD-I) [1]: (a) self-perceived continuous decline in memory compared with the previous normal status within the last 5 years combined with informant report; (b) MMSE and MoCA scores were within the normal range after age-, gender-, and education-adjustment; and (c) CDR score of 0.

Inclusion criteria for aMCI included the following: (a) with memory complaint, confirmed by an informant; (b) objectively impaired memory confirmed by neuropsychological tests; (c) clear-cut history of cognition worsening; (d) can't meet the criteria for dementia according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, revised (DSM-IV-R); and (e) CDR score of 0.5.

d-AD subjects were diagnosed according to the National Institute of Aging-Alzheimer's (NIA-AA) criteria for clinically probable AD [29,30]: (a) meeting criteria for dementia; (b) insidious and gradual onset (not sudden) over more than 6 months; (c) clear-cut history of worsening of cognition; (d) the initial and most prominent cognitive deficits were evident in the performance of amnestic presentation or non-amnestic presentations; and (e) hippocampal atrophy confirmed by structural MRI.

The exclusion criteria applied to all subjects included the following: (a) a history of stroke (Hachinski Ischemic Scale score > 4 points); (b) severe depression (HAMD score > 24 points or the center for Epidemiological Studies Depression Scale score > 21 points); (c) other central nervous system diseases which could cause cognitive decline (e.g., epilepsy, brain tumors, Parkinson's disease, or encephalitis); (d) systemic diseases, which could cause cognitive impairment (e.g., anthracemia, syphilis, thyroid dysfunction, or severe anemia, or HIV); (e) a history of psychosis or congenital mental growth retardation; (f) sever hypopsia or dysacusis; (g) cognitive decline caused by traumatic brain injury; (h) sever end-stage diseases or sever diseases on acute stages; or (i) those who could not complete neuropsychological tests or with contraindication to MRI.

2. Feature extraction

2.1 Processing of DTI

2.1.1 Neuroimaging protocol

All MR scans were performed on a 3.0 Tesla MR system (Siemens Magnetom Trio Tim MRI system, Germany) using a standard head coil. During the entire scanning procedure, cushions and headphones were used to reduce subject motion and scanner noise. Structural DTI data and T1-weighted data was available in all participants.

Diffusion Tensor Imaging. DTI data was collected using an echo planar imaging (EPI) sequence with following parameters for three times: in 31 independent, non-collinear directions of a b-value = 1000 and one additional image with no diffusion weighting (b = 0), slices = 60, TR= 11000 ms, TE =98 ms, flip angle =90°, FOV = 256 mm×232 mm, acquisition matrix= 128×116, and thickness= 2 mm with reversed k-space read-out. The resulting T1-weighted images and cortical models were linearly aligned to the space of the diffusion weighted imagings (DWIs). DWIs as well as resulting tracts were further elastically registered to the T1-weighted image to account for susceptibility artifacts (we assume that the T1-weighted scan serves as a relatively undistorted anatomical reference).

Anatomical T1. In addition, a T1-weighted image was acquired for anatomical reference. T1-weighted MR images were obtained by a 3D magnetization-prepared rapid gradient echo (MPRAGE) with following parameters: Slices = 176, TR = 1900 ms, TE = 2 ms, inversion time (TI) = 900 ms, flip angle=9°, field of view (FOV) = 224×256, acquisition matrix = 448×512, no gap, and thickness = 1.0 mm [31].

2.1.2 Image Pre-processing

Image preprocessing steps were performed using the PANDA toolbox (http://www.nitrc.org/projects/panda) based on FSL 5.0 for all DTI images (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki) and Diffusion Toolkit (http://www.trackvis.org/dtk/), including motion and eddy current corrections[32]. Fractional anisotropy (FA) of each voxel was computed, and a high FA value indicating a high level of directionally restricted diffusion of water molecules.

Affine transformation was used to co-register FA images in native space to its corresponding T1-weighted images. Structural images were then non-linearly registered to the ICBM152 template. Based on the above two steps, an inverse warping transformation from the standard

space to the native dMRI space can be obtained. Automated anatomical labeling (AAL) atlas in the standard space were then inversely warped back to individual native space by applying this inverse transformation. This parcellation divided the cortical surface into 90 regions (45 per hemisphere). See Table 2 for the name of the regions and their corresponding abbreviations. The resulting inverse deformation map (T^{-1}) for each subject was then applied to warp the AAL template to the DTI native space of each subject, using nearest neighbor interpolation method as each AAL region was defined as a brain network node.

Finally, for each individual DTI data set, WM pathways were reconstructed and referred to as fibers or tracts using streamline tractography. The Fiber Assignment by Continuous Tracking (FACT) algorithm was used [33,34]. In the brain mask, 8 seeds followed the main diffusion direction from voxel to voxel; and a streamline was terminated when it reached a voxel with a FA value lower than 0.1 (reflecting low levels of preferred diffusion, often grey matter voxels), or when the streamline exceeded the brain mask (i.e. grey and white matter voxels), or when the trajectory of the streamline made a turn sharper than 45 degrees[35]. Streamlines longer than 15 mm were considered for further analysis.

2.2 Processing the Resting-State fMRI (rs-fMRI)

Post-processing of the rs-fMRI images, such as slice timing correction and head-motion correction were performed using the Statistical Parametric Mapping software package (SPM8, http://www.fil.ion.ucl.ac.uk.spm). To ensure magnetization equilibrium, the first 10 acquired fMRI images of each subject were discarded. The remaining 140 images were first corrected for the acquisition time delay among different slices before they were realigned to the first volume of the remaining images for head-motion correction. We hypothesize that the variability of BOLD signal of GM regions is sensitive in delineating the alteration of connectivity patterns, caused by pathological attacks of MCI. Removing signal from the ventricles and WM is motivated by the fact that these regions contain a relatively high proportion of noise caused by the cardiac and respiratory cycles[36]. Accordingly, we first segmented the T1-weighted image of each subject into GM, WM and CSF. For each subject, the GM was then used to mask the fMRI images. This procedure eliminated the possible contribution from WM and CSF in the re-fMRI time series.

The first scan of fMRI time series was coregistered to the T1-weighted image of same subject. The estimated transformation was then applied to other fMRI scans of the same subject. Deformation fields were estimated by warping the Automated Anatomical Labeling (AAL) [28] template (T1-weighted image) to the subject T1-weighted images using a deformable registration method called HAMMER[37]. The brain space of each subject was then parcellated into 90 ROIs by warping the AAL region masks to the subject space using the estimated deformation fields. For each subject, the mean time series of each individual ROI was computed by averaging the GM-masked fMRI time series over all voxels in that particular ROI.

One crucial step in rs-fMRI analysis is temporal band-pass filtering. The frequency interval of band-pass filtering varies and depends on the application, but is normally within the interval of [0.01-0.10Hz] since the fMRI dynamics of neuronal activities are most salient within this frequency interval. It provides a reasonable trade-off between avoiding the physiological noise associated with higher frequency oscillations[38] and the measurement error associated with estimating very low frequency correlations from limited time series[39].

The analysis of rs-fMRI is normally performed on full spectrum of the filtered signals - a relatively global analysis which might not be sensitive enough to delineate complex yet subtle pathological patterns related to the neurological disease. Such global analysis on BOLD signal might cause local, subtle temporal changes to be averaged out, and thus deteriorate classification performance. A relatively local analysis, which is more sensitive to BOLD signal changes, is hence required.

In order to extract complex, yet subtle pathological influences of MCI, we employed a multi-spectrum characterization of the regional mean time series, which utilizes multiple frequency sub-bands, in contrast to the conventional full-spectrum description, to construct functional connectivity networks. The GM-masked mean time series of each region was band-pass filtered within frequency interval [$0.025 \leq f \leq 0.100$Hz] before it was decomposed into five distinct, equally divided frequency sub-bands using the Fast Fourier transform (FFT), enabling a relatively frequency specific analysis of the regional mean time series. By using this multi-spectral characterization, a relatively local analysis, which is more sensitive in delineating complex yet subtle pathological patterns related to the neurological disease, can be achieved.

Functional connectivity, which represents interregional correlations in neuronal variability, was measured using pairwise Pearson correlation coefficients between the ROI pairs. Given a set of N random variables, the Pearson correlation matrix is a symmetric matrix in which each off-diagonal element is the correlation coefficient between a pair of variables. We considered the brain regions as a set of nodes and the correlation coefficients as signed weights represented by the edges connecting the nodes. Fisher's r-to-z transformation was applied on the elements of the Pearson correlation matrix to improve the normality of the correlation coefficients as (1) where r is the Pearson correlation coefficient and z is approximately a normal distribution with standard deviation. The functional connectivity networks are represented in the form of z-maps. Examples of the constructed functional connectivity maps for a normal control (NC) and an individual with MCI are shown in the top and bottom rows of Figure 3, respectively.

Hence, a total of 90 features can be obtained from each connectivity map, producing for each subject a pool of ($6 \times 90 = 540$) and ($5 \times 90 = 450$) features for DTI and fMRI modalities, respectively.

3. Feature selection

We thus employed multiple feature selection methods, including (1) lasso[40], (2) dirty model [41], and (3) extended dirty model[42].

We employ a sparse regression method (Liu et al. 2014; Wee et al. 2014a; Zhang et al. 2012b) to deal with small sample size problem. Since the two target responses, i.e., class labels (patients:+1; normal controls: -1) yeCT and clinical score (SRS_TOTAL) yeST are correlated, we apply a multi-task learning algorithm for feature selection. Here, each task is associated with one target response.

3.1 Lasso

Suppose that we have the data set $(\mathbf{x}^i, y_i), i = 1, 2, \ldots, N$, where $\mathbf{x}^i = (x_{i1}, \ldots, x_{ip})^T$ are the predictor variables and y_i are the responses. As a usual practice in regression, we assume

either that the observations are independent or that the y_is are conditionally independent of the x_{ij}s given. We assume that the x_{ij} are standardized so that $\sum_i x_{ij}/N = 0$, $\sum_i x_{ij}^2/N = 1$.

Letting $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$, the lasso estimate $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ is defined by

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \arg\min \left\{ \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\} \qquad \text{subject to } \sum_j |\beta_j| \le t. \quad (1)$$

Here $t \ge 0$ is a turning parameter. Now for all t, the solution for α is $\hat{\alpha} = \bar{y}$. We can assume without loss of generality that $\bar{y} = 0$ and hence omit α.

Computation of the solution to equation (1) is a quadratic programming problem with linear inequality constraints.

The parameter $t \ge 0$ controls the amount of shrinkage that is applied to the estimates. Let $\hat{\beta}_j^o$ be the full least squares estimates and let $t_0 = \sum |\hat{\beta}_j^o|$. Values of $t \le t_0$ will cause shrinkage of the solutions towards 0, and some coefficients may be exactly equal to 0. For example, if $t = t_0/2$, the effect will be roughly similar to finding the best subset of size $p/2$. Note also that the design matrix need not be of full rank.

4. Multiple-Kernel SVM classification

After performing selection of the discriminative and common features (i.e., brain regions) across multiple modalities, we then followed the multi-kernel SVM method proposed in (Zhang et al. 2011) for the classification of AD/MCI/SCD from NCs.

Based on the features obtained from the proposed method, we then compute a linear, poly, and rbf kernel among all subjects for each modality and use the following function to integrate the multiple kernels:

- A linear kernel based SVM classifier based on the LIBSVM library[43] was adopted in the current mixed-kernel SVM classier.
- Individual kernel matrices are obtained from the selected features of each modality before integrated into a single mixed-kernel matrix.
- A linear SVM produces a multi-dimensional hyperplane that optimally separates data in labeled groups (supervised learning).

Advantages of this multivariate method over a univariate method include (1) increased statistical power, and (2) single subject examination applicability, with the capability to process large amounts of dependent voxel data that resemble global brain functioning more accurately[44].

Before we computed the kernel between pairs of feature vectors, we first performed a normalization step on each feature vector to obtain unit norm vector (i.e., $\|x\|_2 = 1$). In fact, linear kernel with this normalization step can be regarded as a "normalized linear kernel".

Before applying the SVM algorithm to the data, all training data was rescaled so that the values of every feature ranged between zero and one.

K-fold and libsvm

To test the superiority of the proposed method, we employed 10-fold cross-validations. Specifically, the dataset was randomly partitioned into 10 non-overlapping subsets; 9 out of the 10 subsets were used for training, whereas the remaining one for testing. To further avoid possible biases during partitioning, we repeated the experiments 10 times. In each experiment, we built one site classifier for the identification of the scanning site where a test MR image was obtained, and also constructed four IQ score estimators, i.e., one multi-kernel SVR for each scanning site. For training the site classifier, the training samples of all the datasets were used together; but for training the IQ score estimators for different scanning sites, only the training samples of the respective dataset were used. Note that the process of training site classifier is independent from that of feature selection and regression models. We used a degree-2 polynomial kernel function for multi-kernel SVR. To determine the model parameters, i.e., $\lambda 1$, $\lambda 2$, and $\lambda 3$, kernel parameters c, p and weights β in multi-kernel SVR, and a sparsity control parameter γ in SMRL, we further divided the training samples for inner cross-validation and then acquired the optimal parameter set that produced the best performance in the inner loop. These parameter values are thenused for the left-out testing samples[45].

Throughout the whole process, we employed LibSVM MATLAB library (version 3.17, Department of Computer Science and Information Engineering, National Taiwan University, Taiwan) [Chang and Lin, 2011] and custom scripts implemented in MATLAB (release 2013a, The MathWorks, Natick, MA).

Finally, the LIBSVM toolbox (Chang and Lin 2011) is also adopted to perform the mixed-kernel SVM defined.

Result

i=5, fold=10	lasso	elastic	dirty model	extended dirty model
AD	96.34	97.2747	97.70	98.37
MCI	93.70	94.044	93.43	93.60
SCD	72.33	73.5	78.67	80.50

In order to further show the superiority of our proposed method, we compare it with other popular feature selection methods including RelieF [46] and Elastic [47].
For fair comparison, we use the same classifier (i.e., multi-kernel SVM) after performing feature selection using RelieF, Elastic Net and our proposed method. Table 6 gives the classification accuracies of different feature selection methods for AD vs. HC, MCI vs. HC and MCI-C vs. MCI-NC, respectively. As we can see from Table 6, our proposed method always achieves the best classification accuracies in all the three classification tasks, compared to RelieF and Elastic Net. In particular, our proposed method exceeds nearly 10 percentage points than other two compared methods in the classification accuracy of MCI-C vs. MCI-NC. This result again validates the efficacy of our proposed method.

Reference

Tales, Andrea, et al. "Subjective Cognitive Decline." *Journal of Alzheimer's Disease* 48.Suppl 1 (2015): S1.

Jessen, Frank, et al. "A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease." *Alzheimer's & Dementia* 10.6 (2014): 844-852.

Glodzik-Sobanska L, Reisberg B, De Santi S, Babb J, S, Pirraglia E, Rich K, E, Brys M, de Leon M, J, Subjective Memory Complaints: Presence, Severity and Future Outcome in Normal Older Subjects. *Dement Geriatr Cogn Disord* 2007;24:177-184

Dufouil, C., Fuhrer, R. and Alpérovitch, A. (2005), Subjective Cognitive Complaints and Cognitive Decline: Consequence or Predictor? The Epidemiology of Vascular Aging Study. *Journal of the American Geriatrics Society*, 53: 616–621. doi:10.1111/j.1532-5415.2005.53209.x

van Oijen M, de Jong FJ, Hofman A, Koudstaal PJ, Breteler MM. Subjective memory complaints, education, and risk of Alzheimer's disease. *Alzheimers Dement* 2007;3:92–7.

Jessen F, Wiese B, Bachmann C, Eifflaender-Gorfer S, Haller F, K€olsch H, et al. Prediction of dementia by subjective memory impairment: effects of severity and temporal association with cognitive impairment. *Arch Gen Psychiatry* 2010;67:414–22.

Reisberg B, Shulman MB, Torossian C, Leng L, Zhu W. Outcome over seven years of healthy adults with and without subjective cognitive impairment. *Alzheimers Dement* 2010;6:11–24.

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild Cognitive Impairment Clinical Characterization and Outcome. *Arch Neurol*. 1999;56(3):303-308. doi:10.1001/archneur.56.3.303

Davatzikos C, Yong F, Wu X, et al. (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging* 29(4):514–523.

Teipel SJ, Kurth J, Krause B, et al. (2015) The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment – Beyond classical regression. *Clin Neuroimaging* 1(4):244–247. in Press.

Xu L, Wu X, Chen K, et al. (2015) Multi-modality sparse representation- based classification for Alzheimer's disease and mild cognitive impairment. *Comput Methods Programs Biomed* 122(2):182–190.

Wee CY, Yap PT, Zhang D, et al. (2012) Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59(3):2045–2056.

Dyrba M, Barkhof F, Fellgiebel A (2015a) Predicting prodromal Alzheimer's disease in subjects with mild cognitive impairment using machine learning classification of multimodal multicenter diffusion-tensor and magnetic resonance imaging data. *J Neuroimaging* 25(5):738–747.

Baron JC, Che´telat G, Desgranges B, et al. (2001) In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage* 14(2):298–309.

Yoko H, Hiroshi M, Kiyotaka N, et al. (2005) Voxel-based morphometry to discriminate early Alzheimer's disease from controls. *Neurosci Lett* 382(3):269–274.

Davatzikos C, Yong F, Wu X, et al. (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging* 29(4):514–523.

Lerch JP, Pruessner J, Zijdenbos AP, et al. (2008) Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol Aging* 29(1):23–30.

Colliot O, Chetelat GM, Desgranges B, et al. (2008) Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology* 248(1):194–201.

Fornito A, Bullmore ET (2010) What can spontaneous fluctuations of the blood oxygenation-level-dependent signal tell us about psychiatric disorders. *Curr Opin Psychiatry* 23(3):239–249.

Zhang Z, Liu Y, Jiang T, et al. (2012) Altered spontaneous activity in Alzheimer's disease and mild cognitive impairment revealed by Regional Homogeneity. *Neuroimage* 59(2):1429–1440.

Zhang Z, Liu Y, Jiang T, et al. (2012) Altered spontaneous activity in Alzheimer's disease and mild cognitive impairment revealed by Regional Homogeneity. *Neuroimage* 59(2):1429–1440.

Chen G, Ward BD, Xie C, et al. (2011) Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology* 259(1):213–221.

Apostolova LG, Hwang KS, Andrawis JP, et al. (2010) 3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects. *Neurobiol Aging* 31(8):1284–1303.

Zhang D, Wang Y, Zhou L, et al. (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55(55):856–867.

Dyrba M, Grothe M, Kirste T, et al. (2015b) Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple Kernel SVM. *Hum Brain Mapp* 36(6):2118–2131.