Psychiatry Research: Neuroimaging 212 (2013) 89–98

Contents lists available at SciVerse ScienceDirect

Psychiatry Research: Neuroimaging

journal homepage: www.elsevier.com/locate/psychresns

Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment

Carlos Aguilar [a,*], Eric Westman [a], J-Sebastian Muehlboeck [b], Patrizia Mecocci [c], Bruno Vellas [d], Magda Tsolaki [e], Iwona Kloszewska [f], Hilkka Soininen [g], Simon Lovestone [b,h], Christian Spenger [i], Andrew Simmons [b,h], Lars-Olof Wahlund [1,a]

[a] Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[b] King's College London, Institute of Psychiatry, London, UK
[c] Institute of Gerantology and Geriatrics, University of Perugia, Perugia, Italy
[d] INSERM U 1027, University of Toulouse, Gerontopole, CHU Toulouse, France
[e] 3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece
[f] Medical University of Lodz, Lodz, Poland
[g] Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[h] NIHR Biomedical Research Centre for Mental Health, London, UK
[i] Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 3 February 2012
Received in revised form
5 November 2012
Accepted 15 November 2012

Keywords:
Multivariate analysis
Machine learning
Magnetic resonance imaging (MRI)
AddNeuroMed
Alzheimer's disease
Mild cognitive impairment

ABSTRACT

Automated structural magnetic resonance imaging (MRI) processing pipelines and different multivariate techniques are gaining popularity for Alzheimer's disease (AD) research. We used four supervised learning methods to classify AD patients and controls (CTL) and to prospectively predict the conversion of mild cognitive impairment (MCI) to AD from baseline MRI data. A total of 345 participants from the AddNeuroMed cohort were included in this study; 116 AD patients, 119 MCI patients and 110 CTL individuals. High resolution sagittal 3D MP-RAGE datasets were acquired and MRI data were processed using FreeSurfer. We explored the classification ability of orthogonal projections to latent structures (OPLS), decision trees (Trees), artificial neural networks (ANN) and support vector machines (SVM). Applying 10-fold cross-validation demonstrated that SVM and OPLS were slightly superior to Trees and ANN, although not statistically significant for distinguishing between AD and CTL. The classification experiments resulted in up to 83% sensitivity and 87% specificity for the best techniques. For the prediction of conversion of MCI patients at baseline to AD at 1-year follow-up, we obtained an accuracy of up to 86%. The value of the multivariate models derived from the classification of AD vs. CTL was shown to be robust and efficient in the identification of MCI converters.

1. Introduction

Alzheimer's disease (AD) is one of the most common forms of neurodegenerative disorders. The disease is related to pathological amyloid depositions and hyper-phosphorylation of structural proteins which leads to progressive loss of cognitive function, synaptic dysfunction and structural changes in the brain. Magnetic resonance imaging (MRI) has been extensively investigated in AD and, consistent with pathology, very early changes have been demonstrated in the hippocampus and entorhinal cortex. However, no imaging measure currently provides a reliable prediction of which patients with mild cognitive impairment (MCI) will rapidly progress to develop AD (O'Brien, 2007; Ries et al., 2008). With MRI it is possible to measure both regional (hippocampus/entorhinal cortex) and global (whole brain) atrophy, which are considered sensitive surrogate markers, capable of quantifying the extent of brain degeneration in dementia (Apostolova et al., 2006). It is possible to obtain multiple volumetric and cortical thickness measures from high resolution MRI by automated segmentation techniques. It has previously been shown that a combination of global and prediction of conversion of MCI to AD as compared with using manual volumetric measures of the hippocampus (still considered to be the gold standard) (Westman et al., 2011c). Further, multivariate analysis using multiple regions in the brain as input gives better accuracy for AD classification and MCI prediction than visual assessment (Scheltens scale for medial temporal lobe atrophy) performed by an experienced radiologist (Westman et al., 2011a). Positive results have also been reported using a whole-brain grey-

* Corresponding author. Tel.: +46 8 585 82 889; mob: +46 76 245 1061; fax: +46 8 585 85 470.
E-mail address: carlos.aguilar@ki.se (C. Aguilar).
[1] for the AddNeuroMed Consortium.

0925-4927/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pscychresns.2012.11.005

matter-based support vector machine (SVM) approach (Kloppel et al., 2008a). A large number of multivariate methods have been introduced in recent years for classifying individual patients with AD using structural MRI (Vemuri et al., 2008; Plant et al., 2010; Kloppel et al., 2008b; Teipel et al., 2007; Davatzikos et al., 2008; Magnin et al., 2009: Fan et al., 2008). However, the lack of studies using multiple methods on the same data has made it difficult to directly compare the results of the different techniques.

In this study we combine multiple morphometric measures derived from an automated pipeline to directly compare different multivariate classifiers. The specific aims were (1) to compare linear and non-linear multivariate methods for the classification of AD vs. cognitively normal controls (CTL) using an automated pipeline; (2) to test the resulting classifiers in predicting AD conversion from the prodromal stage of the disease, MCI; (3) to assess the effect of age, education and APOE genotype in the prediction of AD vs. CTL; and (4) to identify the optimal classifier(s).

2. Materials and methods

2.1. Subjects and inclusion criteria

All participants originated from the AddNeuroMed project, part of InnoMed (Innovative Medicines in Europe), a European Union program designed to make drug discovery more efficient. The project is designed to develop and validate novel surrogate markers in Alzheimer's disease and includes a human neuroimaging component (Simmons et al., 2009, 2011), which includes the collection of MRI data, other biomarkers, clinical and cognitive measures. Data were collected from six different sites across Europe: University of Kuopio, Finland, University of Perugia, Italy, Aristotle University of Thessaloniki, Greece, King's College London, United Kingdom, University of Lodz, Poland, and University of Toulouse, France. MR images from a total of 345 participants were included in this study; 116 AD patients, 119 MCI patients and 110 CTL individuals. Table 1 gives the demographics of the study cohort. All AD and MCI patients were recruited from local memory clinics of the six participating sites while the control individuals (cognitively normal/CTL) were recruited from non-related members of the patient's families, caregiver's relatives or social centres for the elderly. Written consent was obtained where the research participant had capacity, and in those cases where dementia compromised capacity, then assent was obtained from the patient and written consent from a relative, according to local law and process. This study was approved by ethical review boards in each participating country. The inclusion and exclusion criteria were as follows. *Alzheimer's disease*: *Inclusion criteria*: (1) ADRDA/NINCDS and DSM- IV criteria for probable Alzheimer's disease. (2) Mini Mental State Examination score range between 12 and 28. (3) Age 65 years or above. *Exclusion criteria*: (1) Significant neurological or psychiatric illness other than Alzheimer's disease. This would exclude patients with vascular dementia or large infarcts, for example. (2) Significant unstable systematic illness or organ failure. All AD patientss had a Clinical Dementia Rating (CDR) scale score of 0.5 or above.

Mild Cognitive Impairment and Controls: *Inclusion criteria*: (1) Mini Mental State Examination score range between 24 and 30. (2) Geriatric Depression Scale score less than or equal to 5. (3) Age 65 years or above. (4) Medication stable. (5) Good general health. *Exclusion criteria*: (1) Meet the DSM-IV criteria for dementia. (2) Significant neurological or psychiatric illness other than Alzheimer's disease. (3) Significant unstable systematic illness or organ failure. The distinction between MCI and controls was based on two criteria: (1) Subject scores 0 on the Clinical Dementia Rating Scale=control. (2) Subject scores 0.5 on the Clinical Dementia Rating Scale=MCI. For the MCI patients it was preferable that the subject and informant reported occurrence of memory problems.

MCI conversion to AD: MCI conversion was defined when patients met the criteria for MCI at baseline and subsequently the criteria for AD (ADRDA/NINCDS and DSM-IV) at 1-year follow-up as described above.

At baseline, a health interview was administered to all participants following a standardized protocol available at http://www.innomed-addneuromed.com, and data on demographics, medical history, current health status, medication use and family history were collected. For AD and MCI cases, information on the duration and severity of cognitive decline was obtained from the patients or informants through a detailed questionnaire: the presence of memory problems, time of their onset, subsequent course, and investigations performed. All diagnoses, including follow-up diagnoses were made by the local center clinician based on the criteria above.

For all participants, cognition, behaviour, functional status and global severity were assessed and the Mini-Mental State Examination (MMSE) (Folstein et al., 1975) and the Clinical Dementia Rating Scale (CDR) (Hughes et al., 1982) were administered. Evaluation of AD included the Hachinski ischemic scale (Hachinski et al., 1975), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) (Rosen et al., 1984), the Neuropsychiatric Inventory (Cummings et al., 1994) and the Alzheimer's Disease Cooperative Study (ADCS)-Activities of Daily Living Scale (Galasko etal.,1997). Evaluation of MCI and CTL individuals comprised the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Cognitive Battery (Morris et al., 1989) and the Geriatric Depression Scale (GDS) (Yesavage and Sheik, 1986). MRI findings were not known to clinicians, so as not to influence the clinical diagnosis.

2.2. MRI acquisition

Data acquisition for the AddNeuroMed study was designed to be compatible with the Alzheimer Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008). The imaging protocol for both studies included high resolution sagittal 3D T1-weighted magnetization-prepared rapid acquisition with gradient echo (MPRAGE) volume (voxel size $1.1 \times 1.1 \times 1.2$ mm^3) and axial proton density/T2-weighted fast spin echo images. The MPRAGE volume was acquired using a custom pulse sequence specifically designed for the ADNI study to ensure compatibility across scanners. Full brain and skull coverage was required for both of the latter datasets and detailed quality control procedures were carried out on all MR images from both studies according to the AddNeuroMed quality control procedure (Simmons et al., 2009, 2011).

2.3. Regional volume segmentation and cortical thickness parcellation

FreeSurfer (version 4.5.0) was used for image processing and analysis. The pipeline produces regional cortical thickness and volumetric measures. Cortical reconstruction and volumetric segmentation procedures include removal of non-

Table 1
Demographic variables and cognitive scores.

	AddNeuroMed				
	CTL	MCI	AD	MCI-nc	MCI-c
Women/men[a]	51/59	61/58	40/76	53/45	8/13
Age in years[b]	73.0 [53–88]	74.4 [57–89]	75.5 [58–88]	74.7 [64–89]	72.9 [57–81]
Education in years[c]	10.8 [2–25]	8.9 [0–20]	8.0 [0–22]	8.8 [0–18]	9.4 [4–20]
MMSE	29.1 [25–30]	27.1 [24–30]	20.8 [12–28]	27.2 [24–30]	26.6 [24–30]
ADAS1	4.5 [1–9]	6.3 [3–9]	6.7 [2–10]	6.3 [3–9]	6.2 [4–9]
CDR	0	0.5	1.2 [0.5–2]	0.5	0.5

Data are presented as mean [minimum–maximum]. AD=Alzheimer's disease, MCI=mild cognitive impairment, MCI-c=MCI converter, MCI-nc=MCI non-converter, MMSE=Mini Mental State Examination, Alzheimer s Disease Assessment Scale (ADAS)=word list non-learning, CDR=Clinical Dementia Rating.

[a] No differences between groups, Kruskal–Wallis ANOVA $P=0.0503$, df=2 and chi-sq=6.0.

[b] Significantly different between groups, Kruskal–Wallis ANOVA $P=0.0097$, df=2 and chi-sq=9.3; significant difference two-tailed t-test AD vs. CTL: $P=0.004$, df=220 and t-test=2.9; and CTL vs. MCI-nc: $P=0.0483$, df=205 and t-test=2.0.

[c] Significantly different between groups, Kruskal–Wallis ANOVA $P=0.0001$, df=2 and chi-sq=20.6; significant difference two-tailed t-test at $P<0.0017$ for AD vs. CTL (df=213 and t-test=4.8), MCI vs. CTL (df=220 and t-test=3.2) and CTL vs. MCI-nc (df=206 and t-test=3.2).

brain tissue using a hybrid watershed/surface deformation procedure (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, ventricles) (Fischl et al., 2002, 2004; Segonne et al., 2004), intensity normalization (Sled et al., 1998), tessellation of gray matter and white matter boundary, automated topology correction (Fischl et al., 2001; Segonne et al., 2007), and deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000). Once the cortical models are complete, registration to a spherical atlas takes place, which utilizes individual cortical folding patterns to match cortical geometry across individuals (Fischl et al., 1999). This is followed by parcellation of the cerebral cortex into units based on gyral and sulcal structure (Desikan et al., 2006; Fischl et al., 2004). The pipeline generated 68 cortical thickness measures (34 from each hemisphere) and 50 regional volumes. Volumes of white matter hypointensities, optic chiasm, right and left vessel, and right and left choroid plexus were excluded from further analysis.

White matter hypointensities were excluded since most individuals were characterized by zero values. Cortical thickness and volumetric measures from the right and left side were averaged. Average measures were used since this makes the data interpretation easier and prediction accuracy does not significantly change if averaged measures are used. In total 57 MRI measures were used as input variables for multivariate analysis. These segmentation approaches have been used for multivariate classification of Alzheimer's disease and healthy controls (Westman et al., 2010, 2011c, 2011d, 2013), neuropsychological image analysis (Liu et al., 2010c, 2009), imaging-genetic analysis (Furney et al., 2010; Liu et al., 2010a,b) and biomarker discovery (Thambisetty et al., 2008, 2010, 2011, Kiddle et al., 2012).

2.4. Data preprocessing

All volumetric measures from each subject were normalized by the subject's intracranial volume while cortical thickness measures were not normalized (Westman et al., 2012). Preprocessing of each variable was performed using mean centering and unit-variance scaling. Mean centering improves the interpretability of the data, by subtracting the variable average from the data. By doing so, the data set is repositioned around the origin. Large variance variables are more likely to be expressed in modeling than low variance variables. Consequently, unit variance scaling was selected to scale the data appropriately. This scaling method calculates the standard deviation for each variable, and divides it by this number.

2.5. Orthogonal partial least squares to latent structures

Orthogonal partial least squares to latent structures (OPLS) (Bylesjö et al., 2006; Trygg and Wold, 2002; Wiklund et al., 2008) is a supervised multivariate data analysis method included in the software package SIMCA (Umetrics AB, Umea, Sweden). In OPLS, the information related to class separation is found in the first component of the model, the predictive component. The other orthogonal components in the model, if any, relate to variation in the data not connected to class separation. Focusing the information related to class separation on the first component makes data interpretation easier (Wiklund et al., 2008).

Each OPLS model receives a $Q^2(Y)$ value that describes its statistical significance for separating groups. $Q^2(Y)$ values >0.05 are regarded as statistically significant (Eriksson et al., 2006), where

$$Q^2(Y) = 1 - (\mathrm{PRESS}/\mathrm{SSY})$$

PRESS (predictive residual sum of squares)$= \sum (y_{\mathrm{actual}} - y_{\mathrm{predicted}})^2$ and SSY is the total variation of the Y matrix after scaling and mean centering (Eriksson et al., 2006). $Q^2(Y)$ is the fraction of the total variation of the Y_s (expected class values) that can be predicted by a component according to crossvalidation (CV).

2.6. Decision trees

The decision trees algorithm (Trees) used the functional trees implementation, part of the WEKA machine learning software (Hall et al., 2009). Trees are a family of methods designed to find arbitrarily complex Boolean functions from examples. Leaves represent classifications and branches represent conjunctions of attributes that lead to these classifications. A decision tree breaks classification down into a set of choices about each attribute at a time. It can be turned into a set of logical disjunctions, if-then rules. A greedy algorithm builds the tree by choosing the most informative attribute at each step. A smart way to build a tree is to select the attribute which maximizes the expected reduction in information entropy (information gain). In Information Theory, information entropy quantifies the unpredictability of a problem (Shannon, 1948). For a binary classification problem S (two outcomes $\pm$) and one feature F (with values $\{f\} \in \pm$):

$$Gain = H(S) - \sum_{f \in \pm} \frac{|S_f|}{|S|} H(S_f)$$

where the entropy H for a binary problem S is:

$$H(S) = -p_+ log_2 p_+ - p_- log_2 p_-$$

p represents the proportion of positive (negative) samples in the dataset (probability). S_f represents the samples from S that have feature value $F=f$, and $|S_f|$ represents a count of the number of samples in S_f.

2.7. Artificial neural networks

Artificial neural networks (ANN) are a family of powerful algorithms inspired from the nervous system that model how our brain adapts itself and learns. An artificial neuron consists of a set of weighted inputs (synapses), an adder to sum up the input signals (membrane of the cell), and an activation function, which decides whether to fire or not (threshold). A neural network is a nonlinear statistical model. The model is very general since the operation of forming nonlinear functions of linear combinations of inputs generates a large class of models. The model can approximate any continuous function arbitrarily well.

$$f(x) = g\left(\sum_j w_{jk} g\left(\sum_l w_{jl} x_l\right)\right)$$

The units in the hidden layer(s) sum their inputs, add a constant and take a fixed function of the result. The activation functions g are the logistic function (Ripley, 1996).

The ANN used here utilized the multilayer perceptron implementation, part of the WEKA machine learning software (Hall et al., 2009). The initial weights were chosen at random, the decay was set to 0.2, the momentum to 0.9, and the units in the two hidden layers were set to 3 and 5.

2.8. Support vector machines

An alternative approach for inferring regularities from a set of training examples is the support vector (SV) algorithm (Boser et al., 1992; Vapnik, 1995), which chooses points critical for the classification task at hand. The support vectors are elements of the data set that are relevant in separating the two classes from each other. The SV algorithm finds the parameters of the decision function that maximize the margin between training examples and class boundary. The learning principle is based on structural risk minimization (Vapnik, 2006), which addresses the problem of balancing the model's complexity against its success at fitting the data. This is particularly useful when there are few data for training, and there is no prior information. These functions have an alternative dual kernel representation K (Boser et al., 1992) allowing nonlinear classification.

$$f(x) = sgn\left(\sum_i \alpha_i y_i . (\varphi(x), \varphi(x_i)) + b\right) = sgn\left(\sum_i \alpha_i y_i . K(x, x_i) + b\right)$$

The SVM algorithm used was the LIBSVM implementation (Chang and Lin, 2011). The training set was used to optimize the values for the parameters C and γ (gamma) of the nonlinear Gaussian function based on the accuracy of a 5-fold cross-validation. They were found using a grid search routine included in the LIBSVM library following the procedure suggested by Hsu et al. (2010).

2.9. Classification experiments

Multivariate models were created to distinguish between AD and CTL using MRI data from the FreeSurfer image analysis pipeline. For multivariate analysis 57 variables were used (23 volumetric and 34 cortical thickness variables). Four classification experiments were carried out for each multivariate technique: (1) MRI data; (2) MRI and age; (3) MRI and education; (4) MRI and APOE genotype. Models were also created based on the top 10 ranked features by a scoring mechanism known as relief (Kira and Rendell, 1992). These features are: hippocampus, amygdala, entorhinal cortex, inferior temporal and middle temporal gyrus, accumbens area, superior temporal gyrus, parahippocampal gyrus, inferior lateral ventricle and CSF volumes. This resulted in a total of 16 models.

A fifth classification experiment was carried out for just the OPLS and SVM methods which have a special mechanism to treat variables of different nature. We applied a multi-kernel learning approach for SVM (Zhang et al., 2011), and as suggested the kernel weights were optimized by using a grid search (0, 0.1, y, 1) to obtain a weight of 0.9 for the MRI-based kernel and 0.1 for the demographics kernel. The complexity constant C was also optimized for the linear SVM for each case. A hierarchical learning approach was used for OPLS, an option included in the software package SIMCA.

For further validation, MCI patients were divided into converters (MCI-c) and non-converters (MCI-nc) at 1-year follow-up. The AD vs. CTL MRI models from the

four different multivariate techniques were used as training sets. MCI patients were then used as an external test set. MCI patients were predicted as either AD-like or control-like (cognitively normal). This was done to investigate how well the four classifiers could prospectively predict future conversion from MCI to AD from baseline data.

Ten-fold cross-validation was used for all our experiments (Hastie et al., 2009). Cross-validation is a statistical method for validating a predictive model which involves building a number of parallel models. These models differ from each other by leaving out a part of the dataset at a time. The omitted data are then predicted by the newly created model. Ten-fold cross validation means that 1/10th of the data is omitted for each cross-validated round (data are omitted once and only once).

2.10. Statistical comparison

The classification performance was based on 10-fold cross-validation for each classifier. It was calculated as the percentage of correctly classified patients and healthy controls (sensitivity and specificity), the percentage of total correct classifications (accuracy), and the area under the receiver operating characteristic (ROC) curve (AUC), which is 0.5 for completely random predictions and 1.0 for perfect separation. The AUC has been proposed as an indicator of the quality of separation, or more formally, the probability of correctly ranking a (normal, abnormal) pair (Hanley and McNeil, 1982; Metz, 2006). The area under the ROC curve has a correspondence with the Wilcoxon statistic, and this relationship provides a way to approximate the standard error, which in turn enables the comparison of two algorithms based on formal statistical criteria (Hanley and McNeil, 1983; Metz et al., 1998; Metz and Kronman, 1980; Metz and Pan 1999). ROCKIT was used for statistical comparison of the results based on the AUC, a ROC analysis software program developed at the University of Chicago and part of the Metz ROC software. *P* values below 0.05 were considered significant and thus reported, and correction for multiple comparisons was made by Bonferroni correction.

3. Results

The classification performance was based on 10-fold cross-validation for each classifier, calculated as the percentage of correctly classified patients and healthy controls (sensitivity and specificity), the percentage of correct classifications (accuracy), and the area under the ROC curve (AUC), which is 0.5 for completely random predictions and 1.0 for perfect separation.

3.1. General classification performance

Investigating the AD vs. CTL model using MRI data as input, an accuracy of 81.9% (AUC 0.78) for the Trees algorithm was obtained. Similar performance was obtained using ANN (accuracy=84.9% and AUC=0.88), SVM (accuracy=83.6% and AUC=0.91) and OPLS (accuracy=84.5% and AUC=0.91) (Table 2 and Fig. 1). No significant differences in classification performance between the classifiers based on ROC analysis were observed.

As part of the classification model, the weights were obtained of a linear SVM and the loadings of OPLS as a way to visualize the variables of importance useful for classification (Fig. 4). Among the top ranked features observed were the following: hippocampus (single most important), amygdala, entorhinal cortex, parahippocampal gyrus, superior temporal gyrus, inferior lateral ventricle, middle temporal gyrus, insula. The Trees algorithm performs a feature selection as part of the learning process with the top three variables being hippocampus, middle temporal gyrus and amygdala volume. It was not possibleto obtain a similar feature ranking by ANN as there was no straightforward way to determine this.

The classification performances remained non-statistically significantly different when either all MRI-derived features or the top 10 ranked features by relief were used (Table 2). Among the top ranked features common to all techniques (information gain, chi-squared, relief, SVM weights, OPLS loadings), hippocampus, amygdala, entorhinal cortex, inferior lateral ventricles, CSF, inferior, superior and middle temporal gyri and temporal pole were found.

3.2. The effect of APOE, age and education

The addition of any of the demographic variables and APOE genotype to the MRI data had no significant influence on the predictive power of the algorithms (see Fig. 2). However, if *age* was added to the MRI data, the performance for Trees became significantly reduced, when compared to ANN, OPLS or SVM ($P<0.005$). Similarly, if *education* was added to the MRI data, the performance for Trees was significantly reduced, when compared to ANN, OPLS or SVM ($P<0.05$).

Two models (see Table 3) created by treating the variables differently using multi-kernel learning (SVM, accuracy=86.7% and AUC=0.92) and hierarchical learning (OPLS, accuracy=87.6% and AUC=0.92) did not significantly improve the performance when compared to the baseline MRI-only model when corrected for multiple comparisons.

Since no significant differences were found by adding any other feature, only models derived from the MRI data were considered for further analysis.

3.3. MCI predictions

Using the MRI data, robust classifiers were built, differentiating between the AD and the CTL group. These models were then applied to the MCI group to identify patients with a pattern of atrophy similar to the AD group and those similar to the CTL group. These were subsequently compared to those MCI patients who converted to AD at follow up (MCI-c) and those who remained stable (tMCI-nc).

Up to 86% of MCI converters were correctly identified as having an AD-like pattern of atrophy (Table 4 and Figs. 3 and 4). No significant difference in prediction between the classifiers was observed.

4. Discussion

All classifiers (OPLS, SVM, ANN and Trees) showed high sensitivity and specificity to distinguish AD from CTL, ranging between 82–85%. When the AUC values were examined, no statistically significant differences were observed. A high correlation in the way the predictions were made was also observed (data not shown), indicating that the use of one classifier or another would yield very similar results. The use of multivariate automated analyses for differentiation of subgroups of MCI patients (MCI-c vs. MCI-nc) yielded similar predictive values and no significant differences between classifiers. All classifiers correctly identified 80–85% of MCI-c, but they did not perform as well in predicting the MCI-nc.

4.1. General classification performance

Methods like OPLS, SVM and LDA have been utilized with prediction accuracies of distinguishing between AD and CTL ranging from 80 to 90%. This is in agreement with a previous study using the AddNeuroMed cohort (Westman et al., 2011a), where AD and CTL individuals were correctly assigned 84% of the time using a 7-fold cross-validation run with OPLS and FreeSurfer. The OPLS method has also been utilized in the ADNI cohort with the same input measures as in the present study (Westman et al., 2011b), where 86% of the individuals were correctly assigned to the appropriate group. Cuingnet et al. reported that using cortical

Table 2
Classification performance of the AD vs. CTL model.

	AD vs. CTL			
	Sensitivity (%)	Specificty (%)	Accuracy (%)	AUC
Trees				
MRI only	78.5 [70.1–85.0]	85.5 [77.7–90.8]	81.9 [76.3–86.3]	0.78 [0.105]
MRI+age	77.6 [69.2–84.2]	85.5 [77.7–90.8]	81.4 [75.8–86.0]	0.85 [0.026]
MRI+edu	81.0 [73.0–87.1]	81.8 [73.6–87.9]	81.4 [75.8–86.0]	0.85 [0.026]
MRI+apoE	84.5 [76.8–90.0]	80.9 [72.6–87.2]	82.7 [77.3–87.1]	0.86 [0.024]
Top 10	74.1 [65.4–81.2]	89.0 [81.4–93.8]	81.9 [76.3–86.3]	0.82 [0.066]
ANN				
MRI only	80.2 [72.0–88.6]	90.0 [83.0–94.3]	84.9 [79.7–89.0]	0.88 [0.022]
MRI+age	78.5 [70.2–85.0]	90.9 [84.1–95.0]	84.5 [79.2–88.7]	0.90 [0.020]
MRI+edu	81.0 [73.0–87.1]	89.1 [81.9–93.7]	85.0 [79.7–89.0]	0.91 [0.019]
MRI+apoE	84.5 [76.8–90.0]	90.0 [83.0–94.3]	87.2 [82.2–91.0]	0.91 [0.019]
Top 10	81.0 [73.0–87.1]	88.2 [80.8–93.0]	84.5 [79.2–88.7]	0.91 [0.024]
SVM				
MRI only	81.0 [73.0–87.1]	86.4 [78.7–91.6]	83.6 [78.3–87.9]	0.91 [0.018]
MRI+age	83.6 [75.8–89.3]	88.2 [80.8–93.0]	85.8 [80.7–89.8]	0.90 [0.020]
MRI+edu	86.2 [78.8–91.3]	90.0 [83.0–94.3]	88.1 [83.2–91.7]	0.92 [0.018]
MRI+apoE	81.0 [73.0–87.1]	84.6 [76.6–90.1]	82.7 [77.3–87.1]	0.91 [0.019]
Top 10	81.9 [73.9–87.8]	85.5 [77.7–90.8]	83.6 [78.3–87.9]	0.91 [0.058]
OPLS				
MRI only	79.3 [71.1–85.7]	90.0 [83.0–94.3]	84.5 [79.2–88.7]	0.91 [0.018]
MRI+age	81.0 [73.0–87.1]	90.0 [83.0–94.3]	85.4 [80.2–89.4]	0.91 [0.018]
MRI+edu	80.2 [72.0–86.4]	91.8 [85.2–95.6]	85.8 [80.7–89.8]	0.92 [0.018]
MRI+apoE	81.0 [73.0–87.1]	89.1 [81.9–93.7]	84.9 [79.7–89.0]	0.92 [0.018]
Top 10	79.3 [71.1–85.7]	87.3 [79.8–92.3]	83.2 [77.8–87.5]	0.91 [0.020]

Data are presented as mean [confidence interval] performance from a 10-fold cross-validation run with 95% probability assurance, the area AUC is presented as mean [standard deviation] with 95% probability assurance. ANN=artificial neural networks, SVM=support vector machines, OPLS=orthogonal projections to latent structures, age=age at MRI scan (in years), edu=years of education, apoE=APOE e4 genotype, AD=Alzheimer's disease, CTL=cognitive normal. AUC=area under the characteristic receiver operating curve. Top 10=model built on only the 10 top ranked features.

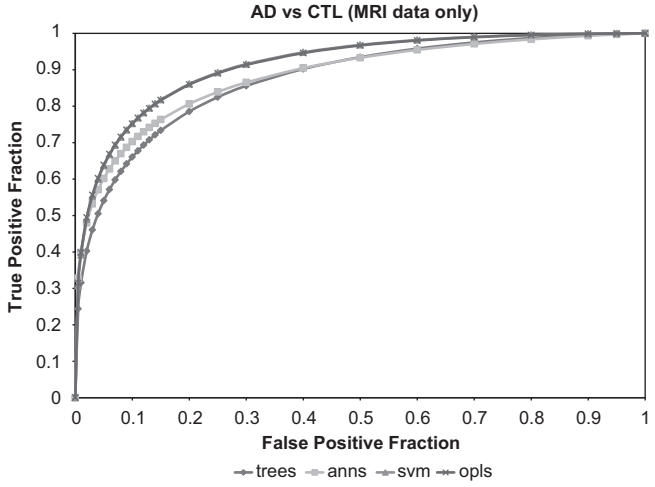

Fig. 1. ROC curve for AD-CTL group using MRI data The curve is calculated with 95% probability assurance. ROC=receiver operating characteristic, ANN=artificial neural networks, SVM=support vector machines, OPLS=orthogonal projections to latent structures, AD=Alzheimer's disease, CTL=healthy control.

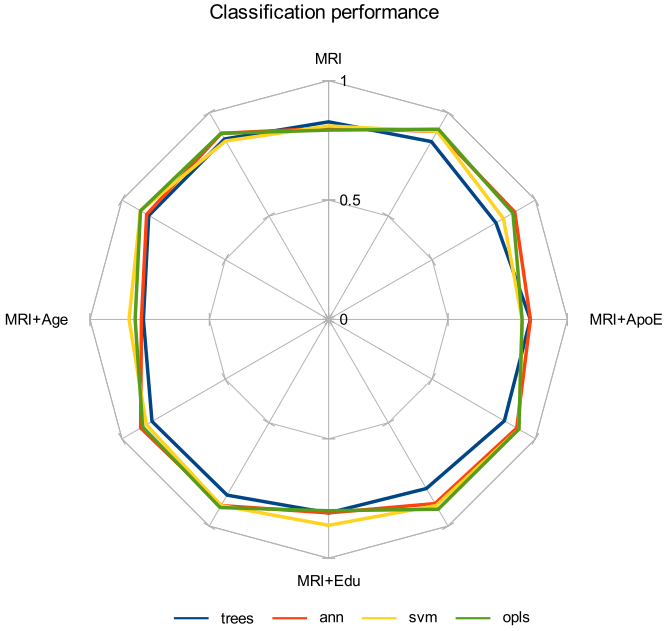

Fig. 2. Radial performance plot for AD-CTL group, MRI data plus an additional variable MRI data only; MRI data+age at MRI scan; MRI data+years of education; MRI data+ApoE. The curve is calculated with 95% probability assurance. ROC=receiver operating characteristic, ANN=artificial neural networks, SVM=support vector machines, OPLS=orthogonal projections to latent structures, AD=Alzheimer's disease, CTL=healthy control.

information derived from FreeSurfer and ADNI data, 85% of the individuals were correctly assigned to the appropriate group. Oliveira et al. reported experiments where 14 AD and 20 CTL individuals from an outpatient local unit were used to train a Gaussian SVM (Oliveira et al., 2010). They reported that using surface-based cortical and volumetric subcortical measurements obtained from FreeSurfer, 88% of the individuals were correctly assigned to the appropriate group applying leave-one-out cross-validation.

Machine learning algorithms have been validated for gene expression analysis (Brown et al., 2000) and for cancer classification (Guyon et al., 2002). Brown et al. compared four machine

Table 3
Classification performance of the AD vs. CTL model adding demographic data (age, education, apoE) to MRI measures.

	AD vs. CTL			
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
SVM				
MRI only	81.0 [73.0–87.1]	86.4 [78.7–91.6]	83.6 [78.3–87.9]	0.91 [0.018]
Multi-kernel	82.8 [74.9–88.6]	90.0 [83.0–94.3]	86.7 [81.7–90.5]	0.92 [0.082]
OPLS				
MRI only	79.3 [71.1–85.7]	90.0 [83.0–94.3]	84.5 [79.2–88.7]	0.91 [0.018]
Hierarchical	84.5 [76.8–90.0]	90.9 [84.1–95.0]	87.6 [82.7–91.3]	0.92 [0.014]

Data are presented as mean [confidence interval] performance from a 10-fold cross-validation run with 95% probability assurance, the area AUC is presented as mean [standard deviation] with 95% probability assurance. SVM=support vector machines, OPLS=orthogonal projections to latent structures, age=age at MRI scan (in years), education=years of education, apoE=APOE e4 genotype, AD=Alzheimer's disease, CTL=cognitive normal. AUC=area under the receiver operating characteristic curve.

Table 4
Predictions obtained by using model I (morphometric variables only) to classify the MCI patients as AD or CTL.

	Trees		ANN		SVM		OPLS	
	AD	CTL	AD	CTL	AD	CTL	AD	CTL
MCI-c	85.7%	14.3%	81.0%	19.1%	85.7%	14.3%	81.0%	19.1%
MCI-nc	51.0%	49.0%	40.8%	59.2%	43.9%	56.1%	31.6%	68.4%
Accuracy	67.4%		70.1%		70.9%		74.7%	
AUC	0.748 [0.076]		0.802 [0.069]		0.807 [0.056]		0.827 [0.056]	

Predictions are presented as the percentage of MCI classified as AD or CTL, MCI-c=MCI converter, MCI-nc=MCI non-converter, the AUC is presented as the mean [standard deviation] with 95% probability assurance. AUC=area under the ROC curve, ROC=receiver operating characteristic, ANN=artificial neural networks, SVM=support vector machines, OPLS=orthogonal projections to latent structures, MCI=mild cognitive impairment, AD=Alzheimer's disease, CTL=healthy control.

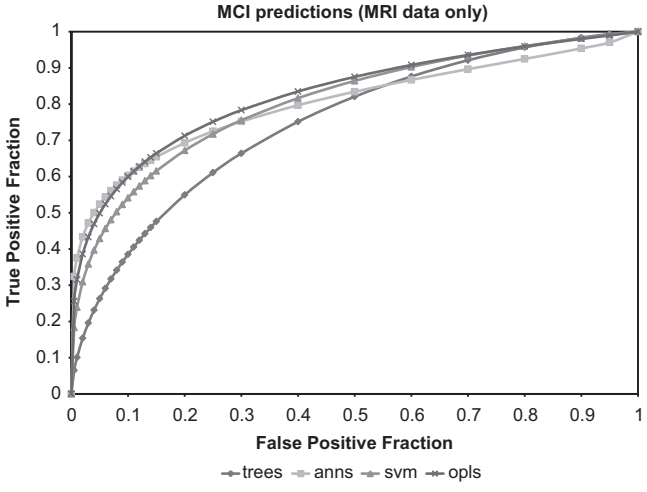

Fig. 3. ROC curves for predictions obtained by using MRI data to classify the MCI patients as AD or CTL The curve is calculated with 95% probability assurance. ROC=receiver operating characteristic, ANN=artificial neural networks, SVM=support vector machines, OPLS=orthogonal projections to latent structures, AD=Alzheimer's disease, CTL=healthy control.

learning algorithms, and found that for gene expression analysis SVM outperformed all other methods investigated. Another study applied and compared eight different classification algorithms for recognition of AD using EEG. They reported that SVM and neural networks performed the best (Lehmann et al., 2007). Our results point in the same direction; if the same input measures are used for this very large dataset, the four different classifiers perform equally well.

The hippocampus was the top ranked relevant feature for classification, agreeing with previous studies which have reported a strong correlation between hippocampal volume and dementia severity (Csernansky et al., 2004). Another important feature was total CSF, which is in accordance with previous results that ventricular expansion significantly differs among AD, CTL and MCI individuals (Evans et al., 2010). Another study (Cuingnet et al., 2011) reported that the main regions of importance for the differentiation of AD vs. CTL were the medial temporal lobe (hippocampus, amygdala, parahippocampal gyrus, cingulum), and the entorhinal cortex, similar to our current findings.

Our results using feature selection showed that the accuracy of the models was not significantly changed using feature selection (Table 2). These models are probably more representative of larger populations than using models with a limited set of features, thus being able to handle biological and disease heterogeneity. We used feature selection by applying data-driven approaches (no prior knowledge) and found that even for Trees, there was no statistically significant improvement in classification.

4.2. MCI predictions

There was a strong association between the structural pattern of atrophy identified in AD and the pattern of atrophy found in MCI converters. However, almost half of MCI non-converters (49–68.4%) were identified as cognitively normal-like, while the others were identified as AD-like. Due to the heterogeneity of the MCI-nc group and short follow-up time, it was not possible to determine the proportion of the MCI-nc group who would subsequently remain stable or convert to AD at some future stage.

It has been proposed that there is a long preclinical phase of AD with no symptoms of cognitive dysfunction but with an ongoing AD pathology (Jack et al., 2010). A recent study has suggested structural changes as detected by MRI may be evident even 10 years before clinical diagnosis of AD (Tondelli et al., 2012). Therefore, it seems possible to detect a stable distribution of brain atrophy suggestive of later cognitive decline, with the medial temporal lobes of particular relevance in the case of MCI and AD. This could potentially explain why so many MCI-nc were identified as AD-like.

4.3. The effect of APOE, age and education

The results of this study show no significant increase/decrease in performance when APOE, age or education was added.

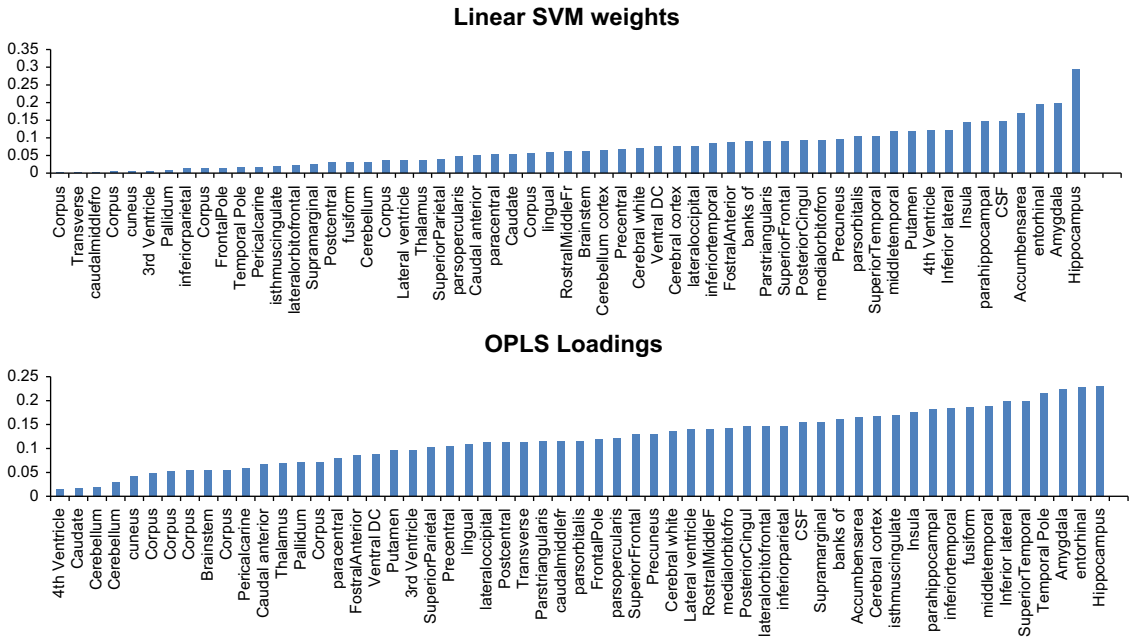

Fig. 4. Scores assigned to MRI-derived measures according to their importance for classification of AD vs. CTL subjects. Fifty-seven MRI measures derived from FreeSurfer (volumes and average cortical thickness) were ranked by taking the absolute value of the weights obtained by a model derived from a linear SVM classifier, and the absolute value of the loading scores assigned by a model derived from an OPLS classifier.

APOE (Lambert and Amouyel, 2011; Corder et al., 1993, 1996; Sun et al., 2012; Parsaik et al., 2012; Burggren et al., 2008; Liu et al., 2010a), education (Stern, 2002; Plassman et al., 1995; Scarmeas and Stern, 2004; Qui et al., 2001; Ngandu et al., 2007; Liu et al., 2012; Munoz et al., 2000; Zhang et al., 2010) and age (Blessed et al., 1968; Savva et al., 2009; Zhang et al., 2010; Holland et al., 2012; Selkoe et al., 2012) are important factors to consider for AD, though they provided no significant additional information to the multivariate models. Potential explanations could be that multivariate analysis is robust to the addition of demographic variables and APOE genotype, that the relationship between brain atrophy and these factors is more complex than the model allows for, or that the overall pattern of atrophy is a strong predictor of AD in and of itself (Kloppel et al., 2008b; Misra et al., 2009; Vemuri et al., 2009, Westman et al., 2011b), without the need for additional information from APOE, age and education.

4.4. Conclusion

Multivariate methods such as Trees, ANN, SVM and OPLS provide a promising way to separate groups based on the simultaneous contribution of all the input variables. They have the advantage of requiring few statistical considerations to obtain good generalization and good estimates of confidence (when used in combination with methods such as cross-validation). This study demonstrates how combining several measures of volume and cortical thickness of different regions of the brain can differentiate between AD patients and healthy controls with high sensitivity and specificity independently of the classifier used. Moreover, high levels of accuracy in the identification of MCI-c as AD-like were reproduced. The results suggest that the different classifiers perform equally well in separating AD from CTL and in predicting MCI conversion. Replication of our results in an independent sample with longer follow-up period would be interesting to validate our results. One drawback of the AddNeuroMed dataset is that it is not neuropathologically confirmed. However, it is very difficult to obtain large neuropathologically confirmed datasets in practice. Neuropathology as a gold standard in dementia has additionally also recently been questioned (Scheltens and Rockwood, 2011), with the suggestion that it should be considered as another complementary biomarker rather than as the gold standard.

Future work could consider more advanced machine learning techniques such as ensemble selection, which combines different classifiers in order to obtain considerably more complex decision boundaries, with the drawback that for many ensemble methods it is not possible to score the features individually, or the required amount of data is greater than for simple algorithms.

Acknowledgements

This study was supported by InnoMed, (Innovative Medicines in Europe), an Integrated Project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5, Life Sciences, Genomics and Biotechnology for Health.

Thanks are also due to the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro), Swedish Brain Power), the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, Loo och Hans Ostermans stiftelse för medicinsk forskning, Stiftelsen för ålderssjukdomar vid Karolinska Institutet, Karolinska Institutets forskningsbidrag, Axel och Signe Lagermans donationsstiftelse, Health Research Council of Academy of Finland, strategic funding from University of Eastern Finland for UEFBRAIN consortium and Stockholm Medical Image Laboratory and Education (SMILE). AS and SL were supported by funds from Alzheimer Research UK and from the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London.

References

Apostolova, L., Dutton, R., Dinov, I., Hayashi, K., Toga, A., Cummings, J., Thompson, P., 2006. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Archives of Neurology 63, 693–699.

Blessed, G., Tomlinson, B.E., Roth, M., 1968. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. British Journal of Psychiatry 114, 797–811.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Worshop on Computational Learning Theory, 144–152.

Brown, M.P.S., Grundy, W.N., Lin, D., Cristianini, N., Sugnet Jr, M.A., C., Haussler, D., 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences of the United States of America 97, 262–267.

Burggren, A.C., Zeineh, M.M., Ekstrom, A.D., Braskie, M.N., Thompson, P.M., Small, G.W., Bookheimer, S.Y., 2008. Reduced cortical thickness in hippocampal subregions among cognitively normal apolipoprotein E e4 carriers. Neuroimage 41, 1177–1183.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J., Holmes, E., Trygg, J., 2006. Opls discriminant analysis: combining the strengths of pls-da and simca classication. Journal of Chemometrics 20, 341–351.

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 1–27.

Corder, E.H., Lannfelt, L., Viitanen, M., Corder, L.S., Manton, K.G., Winblad, B., Basun, H., 1996. Apolipoprotein E genotype determines survival in the oldest ld (85 years or older) who have good cognition. Archives of Neurology 53, 418–422.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A., 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261, 921–923.

Csernansky, J.G., Hamstra, J., Wang, L., McKeel, D., Price, J.L., Gado, M., Morris, J.C., 2004. Correlations between antemortem hippocampal volume and postmortem neuropathology in AD subjects. Alzheimer Disease & Associated Disorders 4, 190–195.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., 2011. and The Alzheimer's Diesease Neuroimaging Initiative, 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56, 766–781.

Cummings, J.L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D.A., Gornbein, J., 1994. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 44, 2308–2314.

Dale, A., Fischl, B., Sereno, M., 1999. Cortical surface-based analysis* 1:: I. segmentation and surface reconstruction. Neuroimage 9, 179–194.

Dale, A.M., Sereno, M.I., 1993. Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. Journal of Cognitive Neuroscience 5, 162–176.

Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M., 2008. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiology of Aging 29, 514–523.

Desikan, R., Segonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., Killiany, R., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J.C.W., Wold, S., 2006. Multi- and megavariate data analysis, second ed. Umetrics AB, Stockholm.

Evans, M.C., Barnes, J., Nielsen, C., Kim, L.G., Clegg, S.L., Blair, M., Leung, L.K., Douiri, A., Boyes, R.G., Ourselin, S., Fox, N.C., 2010. and the Alzheimer's disease Neuroimaging Initiative, 2010. Volumes changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. European Radiology 20, 674–682.

Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., et al., 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39, 1731–1743.

Fischl, B., Dale, A., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences of the United States of America 97, 11050–11055.

Fischl, B., Liu, A., Dale, A., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Transactions on Medical Imaging 20, 70–80.

Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A., 2002. Whole brain segmentation:: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fischl, B., Salat, D., van der Kouwe, A., Makris, N., Segonne, F., Quinn, B., Dale, A., 2004. Sequence-independent segmentation of magnetic resonance images. NeuroImage 23, S69–S84.

Fischl, B., Sereno, M., Tootell, R., Dale, A., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 12, 189–198.

Furney, S., Simmons, A., Breen, G., Pedroso, I., Lunnon, K., Proitsi, P., Hodges, A., Powell, J., Wahlund, L., 2010. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. Molecular Psychiatry 16, 1130–1138.

Galasko, D., Bennett, D., Sano, M., Ernesto, C., Thomas, R., Grundman, M., Ferris, S., 1997. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's disease cooperative study. Alzheimer Disease & Associated Disorders 11 (Suppl 2), S33–39.

Guyon, I., Weston., J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422.

Hachinski, V.C., Iliff, L.D., Zilhka, E., Du Boulay, G.H., McAllister, V.L., Marshall, J., Russell, R.W., Symon, L., 1975. Cerebral blood flow in dementia. Archives of Neurology 32, 632–637.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explorations 11 (1).

Hanley, J., McNeil, B., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148, 839–843.

Hanley, J., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 143, 29–36.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Model Assessment and Selection. The Elements of Statistical Learning. Springer, New York, pp. 219-259.

Holland, D., Desikan, R.S., Dale, A.M., McEvoy, L.K., 2012. for the Alzheimer's disease Neuroimaging Initiative, 2012. Rates of decline in Alzheimer's disease decrease with age. PLoS ONE 7 (8), e4325.

Hsu, C.W., Chang, C.C., Lin, C.J. A Practical Guide to Support Vector Classification. National Taiwan University, 2010.

Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., Martin, R.L., 1982. A new clinical scale for the staging of dementia. The British Journal of Psychiatry 140, 566–572.

Jack Jr, C.R., Bernstein, M., Fox, N., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P., Whitwell, J., Ward, C., Dale, A., Felmlee, J., Gunter, J., Hill, D., Killiany, R., Schu, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C., Krueger, G., Ward, H., Metzger, G., Scott, K., Mallozzi, R., Blezek, D., Levy, J., Debbins, J., Fleisher, A., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27, 685–691.

Jack Jr, C.R., Knopma, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurlogy 9, 119–128.

Kiddle, S.J., Thambisetty, M., Simmons, A., Riddoch-Contreras, J., Hye, A., Westman, E., Pike, I., Ward, M., Johnston, C., Lupton, M.K., Lunnon, K., Soininen, H., Klosweska, I., Tsolaki, M., Vellas, B., Mecocci, P., Lovestone, S., Newhouse, S., Dobson, R., 2012. Plasma based markers of [11C]-PiB brain amyloid burden. PLoS One 7 (9), e44260.

Kira, K., Rendell, L.A., 1992. The feature selection problem: traditional methods and a new algorithm. Proceedings of the 10th National Conference on Artificial Intelligence, 129–134.

Kloppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack Jr, C.R., Ashburner, J., Frackowiak, R.S.J., 2008a. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. Brain 131, 2969–2974.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr, C.R., Ashburner, J., Frackowiak, R.S.J., 2008b. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.

Lambert, J.C., Amouyel, P., 2011. Genetics of Alzheimer's disease: new evidence for an old hypothesis? Current Opinion in Genetics and Development 21, 295–301.

Lehmann, C., Koenig, T., Jelic, V., Prichep, L., John, R.E., Wahlund, L.O., Dodge, Y., Dierks, T., 2007. Application and comparison of classication algorithms for recognition of Alzheimer's disease in electrical brain activity (eeg). Journal of Neuroscience Methods 161, 342–350.

Liu, Y., Paajanen, T., Westman, E., Wahlund, L., Simmons, A., Tunnard, C., Sobow, T., Proitsi, P., Powell, J., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010a. AddNeuroMed Consortium, 2010. Effect of APOE4 allele on cortical thicknesses and volumes: the AddNeuroMed study. Journal of Alzheimer's Disease 21, 947–966.

Liu, Y., Paajanen, T., Westman, E., Wahlund, L., Simmons, A., Tunnard, C., Sobow, T., Proitsi, P., Powell, J., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010b. AddNeuroMed Consortium, 2010. APOE2 allele is associated with larger regional cortical thicknesses and volumes. Dementia and Geriatric Cognitive Disorders 30, 229–237.

Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010c. AddNeuroMed Consortium, 2010. Analysis of regional mri volumes and thicknesses as predictors of conversion from mild cognitive impairment to Alzheimer's disease. Neurobiology of Aging 31, 1375–1385.

Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2009. the AddNeuroMed Consortium, 2009. Combination analysis of neuropsychological tests and structural MRI measures

in diferentiating AD, MCI and control groups-The AddNeuroMed study. Neurobiology of Aging 32, 1198–1206.
Liu, Y., Julkunen, V., Paajanen, T., Westman, E., Wahlund, L.O., Aitken, A., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Spenger, C., Lovestone, S., Simmons, A., Soininen, H., 2012. for the AddNeuroMed Consortium, 2012. Education increases brain reserve in AD, MCI, and healthy controls—evidence from regional cortical thickness and volume measures. Neuroradiology 54, 929–938.
Magnin, B., Mesrob, L., Kinkingnehun, S., Pelegrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S., Benali, H., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73–83.
Metz, C., 2006. Receiver operating characteristic (ROC) analysis: a tool for quantitative evaluation of observer performance and imaging systems. Journal of the American College of Radiology 3, 413–422.
Metz, C., Herman, B., Roe, C., 1998. Statistical comparison of two roc-curve estimates obtained from partially-paired datasets. Medical Decision Making 18, 110–121.
Metz, C., Kronman, H., 1980. Statistical signicance tests for binormal roc-curves. Journal of Mathematical Psychology 22, 218–243.
Metz, C., Pan, X., 1999. Proper binormal roc curves: theory and maximum-likelihood estimation. Journal of Mathematical Psychology 43, 1–33.
Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage 44, 1415–1422.
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, 1159–1165.
Munoz, D.G., Ganapathy, G.R., Eliasziw, M., Hachinski, V., 2000. Educational Attainment and Socioeconomic Status of Patients with Autopsy-Confirmed Alzheimer's disease. Archives of Neurology 57, 85–89.
Ngandu, T., Von Strauss, E., Helkala, E.L., Winbald, B., Nissinen, A., Tuomelehto, J., Soinen, H., Kivipelto, M., 2007. Education and dementia What lies behind the association? Neurology 69, 1442–1450.
O'Brien, J.T., 2007. Role of imaging techniques in the diagnosis of dementia. British Journal of Radiology 80 (Spec No 2), S71–S77.
Oliveira Jr, P.P.M., Nitrini, R., Busatto, G., Buchpigel, C., Sato, J.R., Amaro Jr, E., 2010. Use of SVM methods with surface-based cortical and volume subcortical measurements to detect Alzheimer's disease. Journal of Alzheimer's Disease 19, 1263–1272.
Parsaik, A.K., Lapid, M.I., Rummans, T.A., Cha, R.H., Boeve, B.F., Pankratz, V.S., Tangalos, E.G., Petersen R.C., 2012. ApoE and quality of life in nonagenarians. Journal of the American Medical Directors Association 13 (8), 704–707.
Plant, C., Teipel, S.J., Oswald, A., Bohm, C., Meindl, T., Mouraou-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of disease. NeuroImage 50, 162–174.
Plassman, B.L., Welsh, K.A., Helms, B.S., Brandt, J., Page, W.F., Breitner, J.C.S., 1995. Intelligence and education as predictors of cognitive state in late life: a 50-year follow-up. Neurology 45, 1446–1450.
Qui, C., Bäckman, L., Winblad, B., Agüero-Torres, H., Fratiglioni, L., 2001. The influence of education on clinically diagnosed dementia incidence and mortality data from the Kungsholmen project. Archives of Neurology 58, 2034–2039.
Ries, M.L., Carlsson, C.M., Rowley, H.A., Sager, M.A., Gleason, C.E., Asthana, S., Johnson, S.C., 2008. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. Journal of the American Geriatrics Society 56, 920–934.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.
Rosen, W.G., Mohs, R.C., Davis, K.L., 1984. A new rating scale for Alzheimer's disease. American Journal of Psychiatry 141, 1356–1364.
Savva, G.M., Wharton, S.B., Ince, P.G., Forster, G., Matthews, F.E., Brayne, C., 2009. Age, neuropathology, and dementia. The New England Journal of Medicine 360, 2302–2309.
Scarmeas, N., Stern, Y., 2004. Cognitive reserve: implications for diagnosis and prevention of Alzheimer's disease. Current Neurology and Neuroscience Reports 4, 374–380.
Scheltens, P., Rockwood, K., 2011. How golden is the gold standard of neuropathology in dementia? Alzheimer's and Dementia 7, 486–489.
Segonne, F., Dale, A., Busa, E., Glessner, M., Salat, D., Hahn, H., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.
Segonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Transactions on Medical Imaging 26, 518–529.
Selkoe, D., Mandelkow, E., Holtzman, D., 2012. Deciphering Alzheimer's disease. Cold Spring Harb Perspective in Medicine 2, 1–8.
Shannon, C., 1948. A mathematical theory of communication, Part I. The Bell System Technical Journal 27, 379–423.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2011. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: experience from the rst 24 months. International Journal of Geriatric Psychiatry 26, 75–82.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2009. AddNeuroMed Consortium, 2009. MRI measures of Alzheimer's disease and the AddNeuroMed Study. Annals of the New York Academy of Sciences 1180, 47–55.
Sled, J., Zijdenbos, A., Evans, A., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97.
Sun, X., Nicholar, J., Walker, A., Wagner, M.T., Bachman, D., 2012. APOE genotype in the diagnosis of Alzheimer's disease in patients with cognitive impairment. American Journal of Alzheimer's Disease and Other Dementias 27, 315–320.
Stern, Y., 2002. What is cognitive reserve? Theory and research application of the reserve concept. Journal of the International Neuropyschological Society 8, 448–460.
Teipel, S.J., Born, C., Eewers, M., Bokde, A.L.W., Reiser, M.F., Muller, H.J., Hampel, H., 2007. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 38, 13–24.
Thambisetty, M., Hye, A., Foy, C., Daly, E., Glover, A., Cooper, A., Simmons, A., Murphy, D., Lovestone, S., 2008. Proteome-based identication of plasma proteins associated with hippocampal metabolism in early Alzheimer's disease. Journal of Neurology 255, 1712–1720.
Thambisetty, M., Simmons, A., Velayudhan, L., Hye, A., Campbell, J., Zhang, Y., Wahlund, L., Westman, E., Kinsey, A., Guntert, A., Proitsi, P., Powell, J., Causevic, M., Killick, R., Lunnon, K., Lynham, S., Broadstock, M., Choudhry, F., Howlett, D., Williams, R., Sharp, S., Mitchelmore, C., Tunnard, C., Leung, R., Foy, C., O'Brien, D., Breen, G., Furney, S., Ward, M., K loszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Hodges, A., Murphy, D., Parkins, S., Richardson, J., Resnick, S., Ferrucci, L., Wong, D., Zhou, Y., Muehlboeck, S., Evans, A., Francis, P., Spenger, C., Lovestone, S., 2010. Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer's disease. Archives of General Psychiatry 67, 739–748.
Thambisetty, M., Simmons, A., Hye, A., Campbell, J., Zhang, Y., Wahlund, L.-O., Kinsey, A., Causevic, M., Killick, R., Broadstock, M., Tunnard, C., Leung, R., Foy, C., O'Brien, D., Prinz, T., Ward, M., Kloszewska, I., Mecozzi, P., Soininen, H., Tsolaki, M., Vellas, B., Murphy, D., Parkins, S., Muehlboeck, S., Evans, A., Francis, P., Spenger, C., Lovestone, S., 2011. for the AddNeuroMed consortium, 2011. Plasma biomarkers of brain atrophy in Alzheimer's disease. PLoS One 6 (12), e28527.
Tondelli, M., Wilcock, G.K., Nichelli, P., De Jager, C.A., Jenkinson, M., Zamboni, G., 2012. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiology of Aging 33 825.e25-825-e36.
Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (OPLS). Journal of Chemometrics 16, 119–128.
Vapnik, V., 2006. Estimation of Dependences Based on Empirical Data, second ed. Springer, New York.
Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, Berlin, Germany.
Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwill, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr, C.R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MRI images: validation studies. NeuroImage 39, 1186–1197.
Vemuri, P., Wiste, H.J., WEigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack Jr, C.R., 2009. MRI and CSF biomarkers in normal, MCI and AD subjects: predicting future clinical change. Neurology 73, 294–301.
Westman, E., Cavallin, L., Muehlboeck, J.S., Zhang, Y., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Spenger, C., Lovestone, S., Simmons, A., Wahlund, L.O., 2011a. Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. PLoS ONE 6, e22506.
Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O., 2011b. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage 58, 818–828.
Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L.O., for the AddNeuroMed consortium, 2011c. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. NeuroImage 54, 1178–1187.
Westman, E., Wahlund, L.O., Foy, C., Poppe, M., Cooper, A., Murphy, D., Spenger, C., Lovestone, A., Simmons, A., 2010. Combining MRI and MRS to distinguish between Alzheimer's disease and healthy controls. Journal of Alzheimer's Disease 22, 171–181.
Westman, E., Wahlund, L.O., Foy, C., Poppe, M., Cooper, A., Murphy, D., Spenger, C., Lovestone, S., Simmons, A., 2011d. Magnetic resonance imaging and magnetic resonance spectroscopy for detection of early Alzheimer's disease. Journal of Alzheimer's Disease 26, 307–319.
Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topography 26 (1), 9–23.
Westman, E., Muehlboeck, J.S., Simmons, A., 2012. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage 62, 229–238.

Wiklund, S., Johansson, E., Sjostrom, L., Mellerowicz, E., Edlund, U., Shockcor, J., Gottfries, J., Moritz, T., Trygg, J., 2008. Visualization of GC/TOF-MS-based metabolomics data for identication of biochemically interesting compounds using OPLS class models. Analytical Chemistry 80, 115–122.

Yesavage, J., Sheik, J., 1986. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clinical Gerontologist 5, 165–173.

Zhang, Y., Qiu, C., Lindberg, O., Bronge, L., Aspelin, P., Backman, L., Fratiglioni, L., Wahlund, L.O., 2010. Acceleration of hippocampal atrophy in a non-demented elderly population: the SNAC-K study. International Psychogeriatrics 22, 14–25.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55, 856–867.