Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging

ARTICLE *in* JOURNAL OF ALZHEIMER'S DISEASE: JAD · APRIL 2014

Impact Factor: 4.15 · DOI: 10.3233/JAD-131928 · Source: PubMed

3 AUTHORS, INCLUDING:

Farshad Falahati
Karolinska Institutet
12 PUBLICATIONS **30** CITATIONS

SEE PROFILE

Andy Simmons
King's College London
498 PUBLICATIONS **20,128** CITATIONS

SEE PROFILE

Journal of Alzheimer's Disease 41 (2014) 685–708
DOI 10.3233/JAD-131928
IOS Press

Review

Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging

Farshad Falahati[a,1], Eric Westman[a,*,1] and Andrew Simmons[b,c]
[a]*Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden*
[b]*King's College London, Institute of Psychiatry, London, UK*
[c]*NIHR Biomedical Research Centre for Mental Health, London, UK*

Handling Associate Editor: Patrizia Mecocci

Accepted 2 March 2014

Abstract. Machine learning algorithms and multivariate data analysis methods have been widely utilized in the field of Alzheimer's disease (AD) research in recent years. Advances in medical imaging and medical image analysis have provided a means to generate and extract valuable neuroimaging information. Automatic classification techniques provide tools to analyze this information and observe inherent disease-related patterns in the data. In particular, these classifiers have been used to discriminate AD patients from healthy control subjects and to predict conversion from mild cognitive impairment to AD. In this paper, recent studies are reviewed that have used machine learning and multivariate analysis in the field of AD research. The main focus is on studies that used structural magnetic resonance imaging (MRI), but studies that included positron emission tomography and cerebrospinal fluid biomarkers in addition to MRI are also considered. A wide variety of materials and methods has been employed in different studies, resulting in a range of different outcomes. Influential factors such as classifiers, feature extraction algorithms, feature selection methods, validation approaches, and cohort properties are reviewed, as well as key MRI-based and multi-modal based studies. Current and future trends are discussed.

Keywords: Alzheimer's disease, cerebrospinal fluid, classification, machine learning, magnetic resonance imaging, mild cognitive impairment, multivariate analysis, positron emission tomography

INTRODUCTION

Alzheimer's disease (AD) has long lacked non-invasive *in vivo* biomarkers. Growing evidence for the use of biomarkers has led to their incorporation into more recent criteria for diagnosis of the disease [1]. Evidence supports abnormal processing of the amyloid-β (Aβ) peptide as a possible initiating event of AD [2]. This leads to a series of abnormal changes in the brain such as formation of Aβ plaques, metabolic alterations, synaptic dysfunction, cell death, brain shrinkage, and finally, cognitive decline. A hypothetical model has been proposed that envisages biomarker-based AD staging occurring in a temporally ordered manner [3], with a further refinement to include the preclinical phase of the

[1]These authors contributed equally to this work.

*Correspondence to: Eric Westman, PhD, Karolinska Institutet, Novum, Plan 5, 141 86 Stockholm, Sweden. Tel.: +46 73 655 5179; Fax: +46 8 517 761 11; E-mail: eric.westman@ki.se.

ISSN 1387-2877/14/$27.50

disease [4]. Biomarkers of brain β-amyloidosis are the first to become abnormal. These can be measured by reductions in cerebrospinal fluid (CSF) $A\beta_{42}$ and increased amyloid positron emission tomography (PET) tracer retention. These changes occur before the start of cognitive decline. Secondly, markers of synaptic dysfunction may become abnormal at a very early stage. Neuronal dysfunction and neurodegeneration then follow. Increased CSF tau and structural magnetic resonance imaging (MRI) measures of cerebral atrophy are biomarkers of neuronal injury and neurodegeneration.

The National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) established the criteria for the clinical diagnosis of AD in 1984 [5]. A first revision of this criterion for research purposes was suggested in 2007 [6]. In 2009, the National Institute on Aging (NIA) and the Alzheimer's Association sponsored a series of advisory meetings to establish a process for revising diagnostic and research criteria for AD since the understanding of AD had greatly evolved since 1984 [1]. In the new proposed diagnostic criteria from 2011, only the five most commonly investigated biomarkers highlighted above are included [7]. These criteria have recently been discussed with the conclusion that updates to the criteria are probably warranted every three to four years to incorporate new knowledge about the pathophysiology and progression of the disease [8].

Advances in neuroimaging have resulted in the use of automated techniques for image analysis, which generate large amounts of data for further analysis, rather than simply considering single structures based on manually delineated regions of interests (ROI) in both MRI and PET studies. It is likely that a single ROI is not sufficient for analysis on its own due to the complexity and heterogeneity of AD. Therefore, many recent neuroimaging studies are based on the assumption that the disease is associated with systematic changes in brain structure or function. Looking at patterns of disease using information from the entire brain and combining the different biomarker modalities has shown promising results [9–12]. To analyze such complex patterns of disease using large amounts of data, there is a need for sophisticated multivariate data analysis and machine learning techniques. These tools provide the opportunity to analyze many variables simultaneously and observe inherent patterns in the data. By doing so, it is possible to separate groups, determine which factors cause the separation, and make predictive models of disease.

Recently, many studies have utilized automatic classification techniques including machine learning and multivariate data analysis methods to discriminate AD patients from healthy control subjects (CTL). One of the advantages of these classifiers is the potential use for detecting AD at the prodromal stages, before clinical manifestation. In particular, these techniques have been used to predict conversion from mild cognitive impairment (MCI) to AD, illustrating the potential to be used in clinical practice in the future. Applying multivariate or machine learning techniques to analyze information from different modalities to look for characteristic biomarker signatures may also prove useful when testing new drug therapies. Many treatment strategies have been explored to prevent and slowdown the disease, with limited success [13]. One potential factor for the failure of drugs in clinical trials may be that the wrong population is targeted. The inclusion criteria for clinical trials have largely been based on a battery of cognitive tests for early episodic memory impairment. This is probably not enough due to the complexity and heterogeneity of the disorder. It is likely that a combination of different biomarkers, reflecting different aspects of the disease is needed to include a more homogenous group. Augmenting this with the results of multivariate or machine learning techniques applied to biomarkers is an interesting and potentially exciting avenue. However, the different multivariate and machine learning techniques need to be carefully tested and validated. The studies that have been performed to date report a range of different accuracies for classification and prediction tasks. Practically, the classification accuracy is influenced by several factors including both methods and cohort properties. Analytical factors such as feature extraction methods, feature selection, classification tools, and the robustness of the validation approaches affect the output performance. Moreover, image quality, the number of subjects, demographics (age, gender, genotype, education, etc.) and clinical diagnosis criteria are also important considerations. Most of these previous studies have used different cohorts, features, and techniques. Therefore, to compare different classifiers based on the resultant accuracy can be a complicated task.

This paper will review the recent literature with regard to multivariate analysis and machine learning in AD research. The focus will be on structural MRI since this is today an integrated part of routine clinical work in many settings and one of the most widely investigated of the AD biomarkers. Results will also be reported from studies which combine structural MRI with PET and CSF biomarkers, since all three

are included in the new diagnostic criteria. The review will focus on techniques for classifying individuals rather than group analysis. Results from group analysis are valuable contributions to the literature in their own right, but cannot be transferred to clinical settings where the clinical question relates to the early diagnosis of an individual at risk. Individual classification analyses [14] by contrast have the potential to be used for the early detection of individuals at risk of developing AD in clinical settings.

SEARCH STRATEGY AND SELECTION CRITERIA

We comprehensively searched Medline and PubMed up to November 20, 2013, for articles in English with the search term "Alzheimer's Disease" combined with "machine learning", "multivariate analysis", "classification", "SVM", "OPLS", "PLS", "LDA", "QDA", "Artificial neural network", "decision tree", "MRI", "PET", and "CSF". We read each relevant paper in full, searched their reference lists, and selected the most relevant on the basis of design, findings, and time of publication. Selection criteria were more than 50 subjects in each group, resulting in a minimum of 100 subjects for multivariate modeling (small studies may not be generalizable to the larger population). The search was limited to biomarkers included in the new proposed diagnostic criteria, structural MRI (diffusion MRI, functional MRI, perfusion MRI, and MR spectroscopy were not included) alone or structural MRI combined with PET and/or CSF. In total, 50 articles met all inclusion criteria and were included in this review.

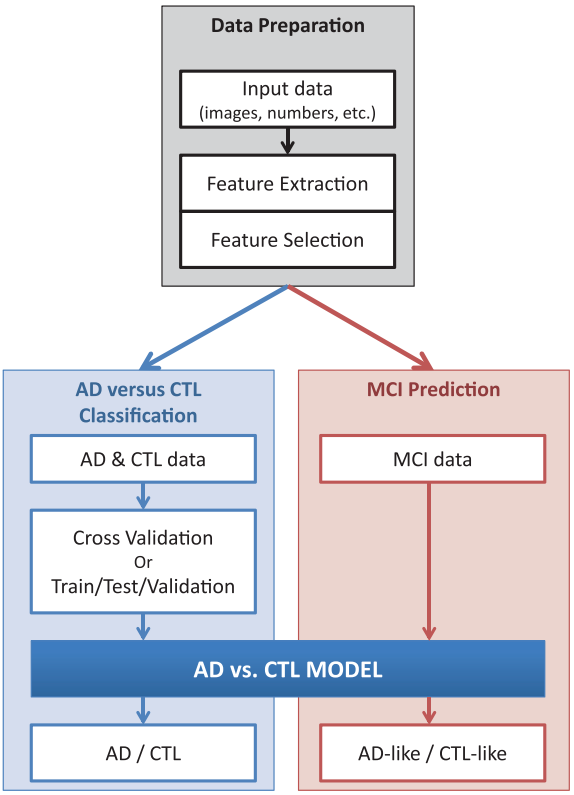

Fig. 1. Illustration of processes in Alzheimer's disease (AD) versus control (CTL) classification and mild cognitive impairment (MCI) prediction. Feature extraction and selection methods are used to obtain the most relevant and discriminative features from the input neuroimaging data (Upper block). Data of AD and CTL subjects are used to create a model for classification of AD and CTL subjects (Left-lower block). MCI subjects are introduced to the created AD versus CTL model in order to be predicted as AD-like and CTL-like (right-lower block).

CLASSIFIERS

Machine learning and multivariate data analysis methods provide the opportunity to process data with high dimensionality and complexity, such as neuroimaging data. Neuroimaging modalities produce extremely high dimensional raw data that can contain inherent patterns related to AD. Machine learning and multivariate data analysis methods are helpful tools for analyzing many variables simultaneously and finding inherent patterns in the neuroimaging data. A wide range of different supervised classifiers has been utilized in the field of AD classification and MCI prediction. Supervised classifiers use prior knowledge about group belonging to learn from a set of training samples. Then the trained classifier can be used to label new unseen samples (Fig. 1).

Support vector machines (SVM) [15, 16] are the most commonly used algorithm in AD research for multivariate classification [9, 17–39]. This method is based on choosing 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 (Fig. 2A). The learning principle is based on structural risk minimization [40], which addresses the problem of balancing the model's complexity against its success at fitting the data. A number of non-linear SVM approaches have also been used such as Kernel-SVM [41] and multi-kernel SVM [12, 42, 43].

The main idea of kernel methods is to map the input data (which is linearly non-separable) into a higher

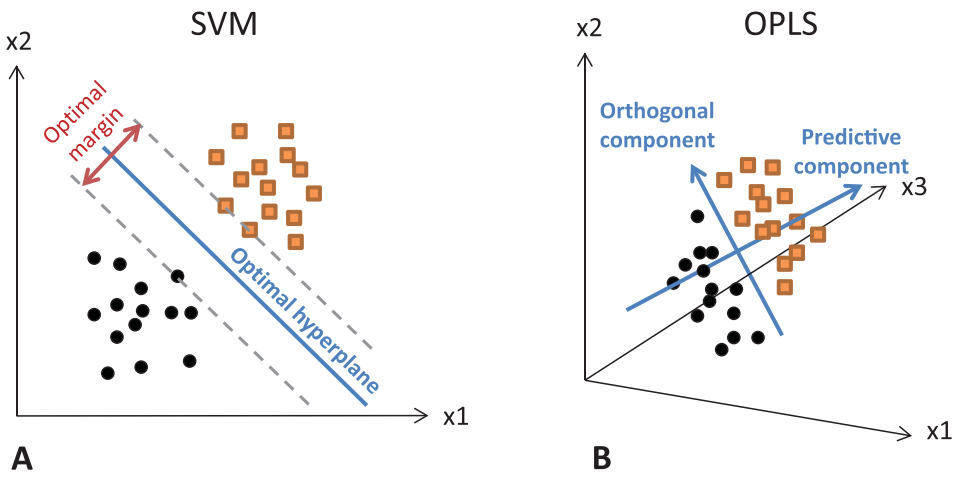

Fig. 2. Simple schematic representation of two common classifiers. A) The support vector machine (SVM) finds a hyperplane that provides the best between classes separation. The optimal hyperplane has the largest distance (optimal margin) to the nearest samples namely support vectors. B) The orthogonal projection to latent structure (OPLS) finds projections that simultaneously maximize the covariance and correlation between data and class label. The predictive component shows between class variation and the orthogonal component shows within class variation.

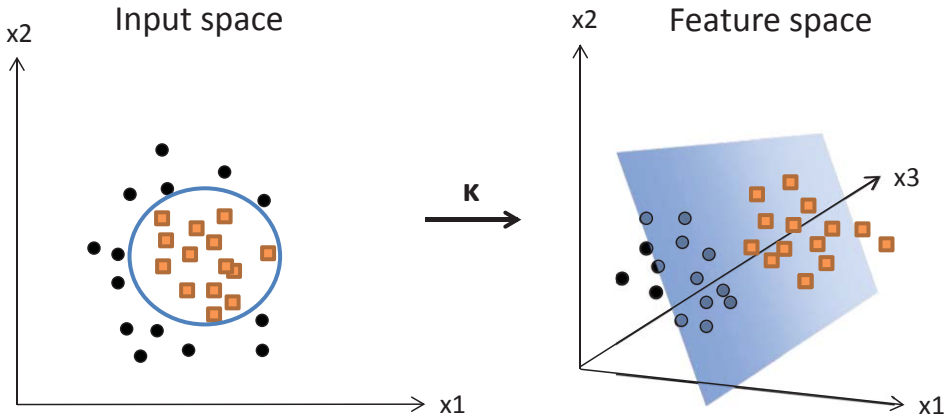

Fig. 3. Schematic representation of the kernel trick. The main idea of kernel methods is to map (by means of a kernel function) the input data, which is linearly non-separable (left) into a higher dimensional feature space, where it is more likely to be linearly separable (right).

dimensional feature space, where it is more likely to be linearly separable (Fig. 3). Explicitly computing the mapping for complex transformation functions can be a limiting factor in terms of computational complexity and requirements. The kernel trick is a well-known way to avoid this limitation by mapping input data into an inner product space without necessity of explicit mapping. The trick is to choose the mapping such that these inner products can be computed within the input space by means of a kernel function. The use of the kernel trick is not limited to the SVM algorithm and other classifiers are also capable of operating with kernels. Multiple kernel learning (MKL) methods [44] are an extension to kernel methods, which instead of one single kernel, construct an optimal combination of base kernels. MKL has been utilized for AD classification and MCI prediction tasks [45, 46].

Orthogonal projection to latent structures (OPLS) [47] is another method which has been used in the AD field [11, 18, 48–52]. This method combines the existing theory of partial least squares (PLS) regression [53, 54] with orthogonal signal correction (OSC) [55, 56]. The PLS method was originally developed for modeling complex data, based on the assumption that there are latent variables which generate the observed data. PLS creates latent vectors (score vectors) by maximizing the covariance between two data sets (independent and dependent variables). PLS can be extended to a regression problem where the dependent variables (say class label) are predicted from a set of independent

variables. The PLS model is negatively affected by systematic variation in the independent variables that is not related to the class labels. The OPLS method is a recent modification of the PLS method to help overcome this problem. Via an inner relationship between the latent variables, OPLS maximizes the covariance between the dependent and the independent variables. 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 (Fig. 2B). Focusing the information related to class separation on the first component makes data interpretation easier [57].

Linear discriminant analysis (LDA) is a well-known statistical classifier that finds a linear combination of features to maximally separate different classes. LDA maximizes the ratio of between-class variance to the within-class variance. Quadratic discriminant analysis (QDA) is a more general version of LDA. Unlike LDA where the covariance matrix is assumed equal for each class, in QDA the class covariance matrices can be different [58, 59]. Both LDA and QDA methods have been utilized for AD classification [17, 29, 36, 60–66].

Other methods used less frequently are artificial neural networks (ANN) [18, 19], decision trees (DT) [18, 67, 68], ensemble methods [30, 37, 69, 70], and regression-based methods [19, 21, 36, 38, 39, 71–73]. ANNs are machine-learning algorithms that are inspired by the central nervous system (brain). An ANN is an interconnected group of nodes, where each node is a computational unit that models a biological neuron. There is a large variety of ANNs, with vast areas of application including classification problems. DTs are a family of methods that use a branching model of decisions and their possible consequences to support decision making. In classification trees [74], the predicted outcome is a discrete category (the class). Ensemble methods are a family of supervised learning algorithms that combine the predictions of multiple basic classifiers into one single classifier. The goal of ensemble methods is to improve accuracy and robustness over a single classifier. Regression analysis is a generic term for many statistical techniques, which discover the relationship between two data sets (independent and dependent variables). Linear regression is the simplest form of regression analysis, which estimates the relationship between independent and dependent variables by a linear model. Logistic regression is another technique, which is used when the dependent variable is categorical, such as diagnostic group.

None of these techniques including SVM and OPLS have specifically been developed for analyzing neuroimaging data but they still all show promising results for AD classification and prediction of conversion from MCI to AD. Table 1 summarizes the studies reviewed in this article according to the classifiers that they employed.

FEATURE EXTRACTION

Feature extraction is of high importance in the field of AD classification. In machine learning, features are a subset of relevant variables that are used as input data to classifiers. Neuroimaging modalities provide comprehensive data on brain structure and function. The main aim of feature extraction techniques is to retrieve and quantify a set of accurate and proper information such as size, shape, volume, etc., from neuroimaging data that can reflect the most relevant disease patterns.

In recent years, many automated and semi-automated feature extraction techniques have been developed and enhanced for analyzing high-resolution structural MRI data. Software package such as FreeSurfer (http://surfer.nmr.mgh.harvard.edu/), FSL (http://fsl.fmrib.ox.ac.uk/), and SPM (http://www.fil.ion.ucl.ac.uk/spm/) provide powerful tools for analyzing MRI data. Feature extraction methods consist of several image-processing and statistical analysis steps that provide a wide range of approaches. Features can vary from single voxels, to ROIs at the level of cortical or subcortical structures, or whole brain. The most common feature extraction techniques include volumetric and thickness measurement methods and morphometry methods. Figure 4 provides examples of different features used.

Cortical/subcortical volumetric and cortical thickness measures have been widely used for classification purposes [18, 48, 49]. Brain morphometry techniques such as voxel-based morphometry (VBM) [75, 76], deformation-based morphometry (DBM) [77, 78], or tensor-based morphometry (TBM) [79] have been used to study structural differences between groups. The main aim of these approaches is to describe and distinguish the macroscopic shape and neuroanatomical configuration differences between different brains [76]. VBM, at its simplest, is a voxel-wise statistical method that identifies differences in the local brain structure (particularly grey matter) between two groups of subjects [76]. Unlike VBM that measures differences in the local composition of brain tissue, TBM and DBM methods detect differences in brain shape

Table 1
A summary of machine learning and multivariate data analysis methods that have been used for AD classification and MCI prediction

Method	No. of articles	Specifications/Properties	References
Support Vector Machines (linear/non-linear)	29	Powerful and accurate classifier Effective in high-dimensional data Good generalization performance Deliver a unique solution Robust to noise/outliers data Computationally expensive	[9, 12, 17–39, 41–43, 45, 46]
Discriminant analysis (LDA/QDA)	9	Well-known classical linear statistical methods Available in a wide range of variations and extensions Simple to implement Suitable for dimensionality reduction in data with high dimensional features Enhance interpretation of between-group differences Optimal for data with Gaussian distribution	[17, 29, 36, 60–66]
Orthogonal Projection to Latent Structures	7	Beneficial for data with large number of dependent/correlated variables Enhance model transparency and improve interpretation Robust to noise/missing data Provide a single predictor component for class separation Detect systematic variation in the data	[11, 18, 48–52]
Decision Trees	3	Simple and fast method Easy to understand and interpret High model transparency Capable of handling noise/outliers/missing data Capable of handling different type of attributes Requires little data preparation	[18, 67, 68]
Artificial Neuronal Networks	2	Powerful nonlinear algorithm No requirement/assumptions on distribution/relationship of input data Capable of accurately handling non-linear and complex patterns Deals well with missing/incomplete data Black box nature (difficult to interpret) Excessive learning time for large neural networks	[18, 19]
Ensemble methods	4	Provide overall higher accuracy than individual classifier Flexible to use different learning algorithms Easy to implement Beneficial for data with high dimensionality and small sample size Robust to noise in data Black box nature (difficult to interpret)	[30, 37, 69, 70]
Regression-based methods	8	Statistical methods with several forms Simple to implement and interpret Flexible on type of input data Include both parametric and non-parametric methods Some assumptions must be considered	[19, 21, 36, 38, 39, 71–73]

LDA, linear discriminant analysis; QDA, quadratic discriminant analysis.

[76]. DBM identifies differences between the positions of brain structures within the brain of subjects, while TBM identifies local shape changes of brain structures between populations [76].

Feature extraction methods based on a single structure have been also developed. Structures in the medial temporal lobe areas, in particular hippocampus and entorhinal cortex, are reported as the first regions to be atrophied in AD patients. Hippocampal volume [17, 80, 81] and shape [20, 41] have been used as input features of classifiers. Furthermore, other less common types of features such as spatial frequency representation of cortical thickness data [61], and spatially normalized tissue density maps [69] have been investigated. The former feature is a map of cortical thickness data transformed into a spatial frequency

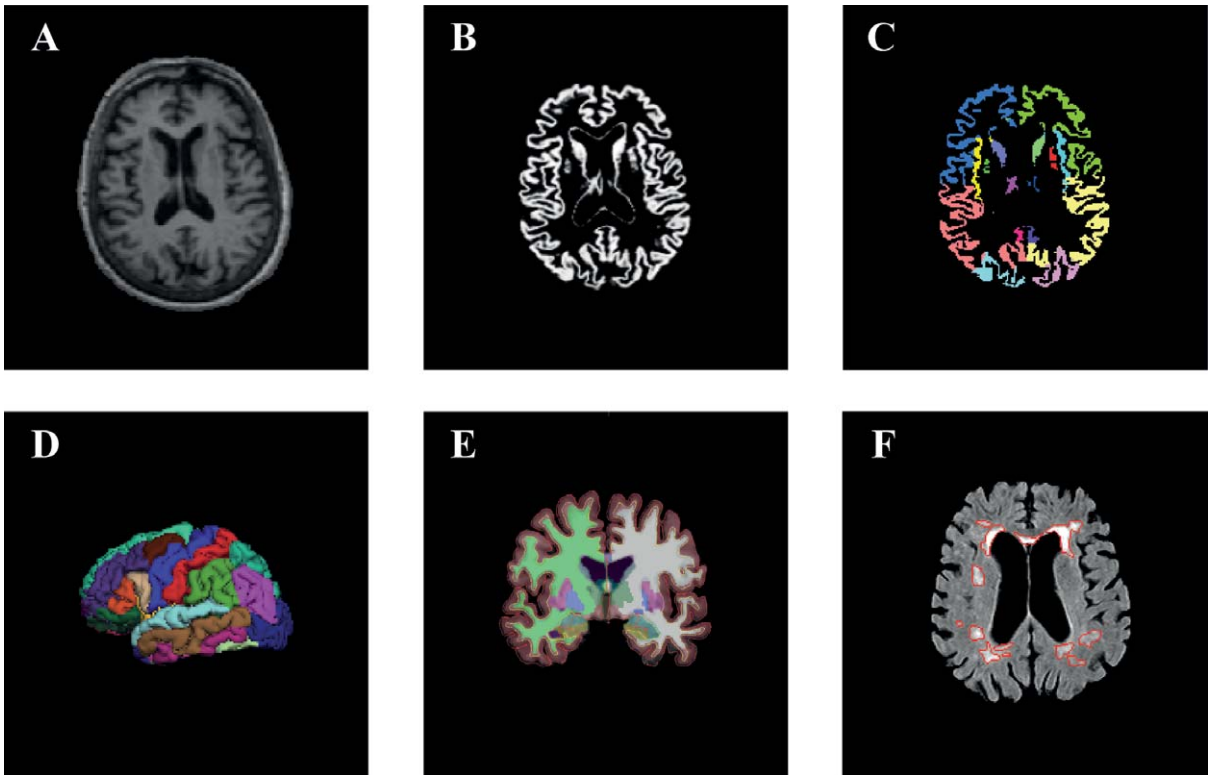

Fig. 4. Examples of input features for classifiers. A) An original axial T1-weighted MR image. B) Grey matter segmentation using SPM. C) Parcellated ROIs after registration image B to an atlas. D) Cortical parcellation using Freesurfer. E) Subcortical and cortical segmentation using Freesufer. F) White matter segmentation using the Cascade software (http://git.io/cascade) [127].

domain, by using a transform function known as manifold harmonic [82]. The latter feature is a map of tissue volumes (grey matter, white matter, and CSF) using a mass-preserving deformable warping algorithm [83].

Feature extraction can directly affect the classification performance. In a recent study, the performance of ten feature extraction methods including five voxel-based methods—three methods based on cortical thickness and two methods ROI-based (hippocampus)—were compared using 509 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort [21]. This comparison resulted in significantly different accuracies. In another study, four different automated feature extraction methods (including hippocampal volume, cortical thickness, TBM, and manifold-based learning) were studied with the results demonstrating that combining the extracted features from different methods improved classification accuracy [17]. Manual measures of hippocampus have long been the gold standard and frequently used in AD studies. However it has been shown that combining both regional and global measures from the entire brain including hippocampal volume using multivariate data analysis results in higher prediction accuracy than using hippocampal volume alone [48].

FEATURE SELECTION

With the help of automated feature extraction methods, it is possible to obtain many features from neuroimaging data. However, because of computational difficulties with high dimensional data (known as the curse of dimensionality), especially in the presence of small numbers of subjects, dealing with many features can be a challenging task, which may result in over-fitting (i.e., the classifier is too closely fit to the small number of subjects) [62]. Using high dimensional data may also lead to noise being introduced to the multivariate models, which can be avoided by using an appropriate feature selection method. Feature selection is an additional helpful stage prior to classification, for dimensionality reduction as well as selecting proper features and omitting improper features. The aim of

feature selection is to select a subset of extracted variables in order to reduce the number of input variables for the classifier. Generally, this step can speed up the classification process and reduce the computational time by decreasing the training and testing time. Moreover, feature selection can reduce model complexity and make model interpretation easier. The effect of feature selection on classification accuracy has been investigated in several studies. Chu et al. compared the predictive accuracies of four different feature selection methods and showed that utilizing a relevant feature selection method based on prior knowledge can improve classification accuracy of an SVM classifier [22]. Further, another study reported that using an SVM-based feature selection method (SVM recursive feature elimination) could improve the performance of MCI prediction [23]. However, some studies have reported that feature selection without prior knowledge did not increase classification accuracy [18, 21, 22]. In addition, it has been reported that feature selection can increase computational time due to adding new hyper-parameters [21].

The main approaches for feature selection are filter-based, wrapper-based, and embedded. Filter-based approaches select the relevant features with regard to the general characteristics of data before introducing features to the classifier. Performing a paired *t*-test to choose the most discriminative features is a common filter-based method [12]. Wrapper-based approaches use a predefined machine learning methods to select features. In SVM classifiers, a specific weight is calculated and assigned to each voxel, which reflects the importance of the voxel in group separation [84]. These weights are potential labels to select important features. In embedded approaches, feature selection is a part of the classifier, e.g., the decision trees method performs feature selection as a part of the training process. Moreover, some more advanced classifiers overcome the curse of dimensionality by utilizing kernel methods [22]. These classifiers take advantage of the kernel trick and transform features from input space to kernel space; therefore the number of subjects determines the input dimensionality of the classifier.

Feature selection may be necessary when using high dimensional data as an input for multivariate data analysis. However, if the methods are robust they should be able to handle noise to a certain extent. Nevertheless, feature selection may make interpretation of the data easier. Using ROIs as input, feature selection may not be needed. ROIs are usually predefined regions of the brain and the numbers generated by any feature extraction method rarely reach levels of high dimensionality.

CROSS-VALIDATION

Cross validation (CV) is a statistical method for evaluating and comparing classifiers. The idea behind CV is to use a part of the dataset to train the classifier, and thereafter use the remaining samples, as a new and unseen set, to test the performance of the classifier. The holdout method is a common approach, where a dataset is randomly divided into two independent training and test subsets. The training subset is only used for the learning procedure and the test subset is used to calculate the performance of the trained classifier. In many classifiers, there are parameters that need to be optimized. In such classifiers, a third subset of samples (the validation subset) is necessary for parameter setting. However, the holdout method depends highly on the distribution of samples in the training dataset, which can easily lead to over-fitting, particularly in the case of a small sample size. The k-fold CV is another method that improves evaluation of classifiers compared to the holdout method. In the k-fold CV, samples are divided into k folds and subsequently k iterations of training and validation are performed, so that each fold is used once and only once for validation. Thus, for each round of CV, the performance of the model can be calculated separately, which decreases the variance of the evaluation. Random sub-sampling and leave-one-out CV are examples of CV. In random sub-sampling, dataset is randomly divided into training and test subsets, and therefore the number of iteration is not limited. Leave-one-out CV is the logical extreme case of k-fold CV, where k is equal to the number of samples in the dataset. Leave-one-out CV may lead to over-fitting of the model [85], but can be useful if only small data sets are available. Leave-one-out [20], 7-fold [48], and 10-fold [42] CV are the most commonly used in the literature. The 10-fold CV has been recommended for the ADNI dataset [86]. The common drawback of CV methods is that the training and testing procedure of the classifier has to be repeated k times, which increases the computation time and cost. However, using fully cross-validated results is recommended to avoid an optimistic bias in classification accuracy [85]. To have a completely independent test set is optimal, but if this is not possible, CV is a good option. However, even if it is possible to use an independent test set, the accuracy of the training model also needs to be validated using CV. If multivariate models are robust the CV method used should not affect the results, but to avoid this potential problem, classifiers would be easier to compare if the same type of CV was used.

COHORTS AND CONFOUNDING FACTORS

Study population is another factor that can affect the classification performance. Baseline characteristics of subjects such as age, gender distribution, genotype, educational level, etc., may have an impact on key biomarkers and are thus influential items in a dataset. Commonly, these characteristics are treated as confounding factors. A confounding factor is defined as a variable in a model that correlates with two factors of interest, i.e., both the dependent and the independent variables. Age, gender, education, and apolipoprotein E (APOE) genotype (a risk factor associated with AD [87]) are examples of confounding factors in the field of AD classification. A common way to deal with confounding factors is to match the subjects of different groups according to the factors. However, it is often not possible to match different clinical groups simultaneously on several parameters. Traditionally, confounding factors are used as covariates in the statistical model to remove their effect from the model [88]. However, this method may not be possible to use in multivariate classifiers. Recently, a method based on minimizing the statistical dependence between the classifier and the confounding factors was introduced for a SVM classifier to control for the effect of confounding factors [89]. In another study, a simple linear detrending method was proposed to correct the effect of age as a confounding factor [90]. This method could be applied prior to statistical evaluation of MRI data using SVM or VBM [90].

Other factors such as the number of subjects and disease severity (degree of impairment) of subjects have been shown to be notable factors [84]. A more severely impaired AD group will show larger structural differences compared to CTL groups, which may lead to higher classification accuracies [84]. Statistically, small datasets may give better results, possibly due to over-fitting, but the results based on large datasets are more reliable and robust. Hardware heterogeneity can also affect the classification performance [91].

Only a few large multicenter cohorts are available in the field of AD. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is the most commonly used cohort. ADNI is a North American-based study that was launched in 2003 and aimed to recruit 800 adults (approximately 200 cognitively normal older individuals, 400 people with MCI, and 200 people with early AD) to participate in the research and be followed for 2-3 years. ADNI subjects aged 55 to 90, and have been recruited from over 50 sites across the U.S. and Canada. Detailed information is available at http://www.adni-info.org. The ADNI study also has two follow up studies, ADNI-GO and ADNI-2.

AddNeuroMed [92, 93] is a part of InnoMed (Innovative Medicines in Europe) [94, 95], which aims to develop and validate novel surrogate markers in AD. The human neuroimaging part of AddNeuroMed includes MRI as an imaging marker of AD. MRI data acquisition was designed to be compatible with the ADNI. Subjects from the AddNeuroMed study were collected from six different sites across Europe. Data from AddNeuroMed is available at http://dataverse.brc.iop.kcl.ac.uk/dvn/dv/ANM.

The Open Access Series of Imaging Studies (OASIS) is a project aiming to freely provide and distribute brain MRI datasets and includes two datasets. The cross-sectional dataset includes MRI data of 416 subjects (young, middle aged, non-demented and demented older adults), aged 18 to 96. The longitudinal dataset includes MRI data of 150 subjects (non-demented and demented older adults), aged 60 to 96. More information is available at http://www.oasis-brains.org/.

Many of the studies in the field have used one of these cohorts, however, some studies have used two cohorts [24] or a combination of cohorts [11, 49]. Although combining different cohorts may result in a more heterogeneous dataset due to different inclusion criteria of subjects, this can lead to the creation of a large and robust model for classification and prediction [11]. There are also several other large cohorts such as The Australian Imaging, Biomarkers and Lifestyle study (http://www.aibl.csiro.au/), DESCRIPA (http://www.descripa.eu/), and the Mayo Clinic Study of Aging [96]. Table 2 shows the cohorts covered by the current review and the number of arti-

Table 2
Cohort representation in multivariate analysis/machine learning papers

Method	No. of articles	References
ADNI	40	[9, 11, 12, 17, 19, 21, 22, 24–26, 28, 30–39, 42, 43, 45, 46, 49–51, 60–64, 66, 68–73]
ANM	6	[11, 18, 41, 48, 49, 52]
CSG-DF	3	[20, 29, 65]
OASIS	1	[24]
ADRC/ADPR	1	[27]
ICBM	1	[29]

ADNI, Alzheimer's Disease Neuroimaging Initiative; ANM, AddNeuroMed Neuroimaging dataset; CSG-DF, IRCSS Centro San Giovanni di Dio Fatebenefratelli; OASIS, The Open Access Series of Imaging Studies; ADRC/ADPR, Mayo Clinic Alzheimer's Disease Research Center/Alzheimer's disease Patient Registry; ICBM, International Consortium for Brain Mapping.

cles that used each cohort. The choice of cohort may or may not affect the results of the classifier [11]. Therefore, it can be difficult to compare results between classifiers using different cohorts.

CLASSIFICATION RESULTS

In this section, five studies that used only MRI data and five studies that used multi-modality data for classification are discussed. The multi-modality studies used PET and/or CSF markers in addition to MRI data. The MRI only studies selected all use data from the ADNI cohort to make the comparison easier to interpret. Studies were selected using the following selection criteria: large numbers of subjects, the study should show results from both AD versus CTL classification and MCI predictions (using the AD versus CTL model to train the model), the studies should use CV for classification and have high classification/prediction accuracies. Finally, the studies were chosen to represent as many different techniques as possible in addition to the above criteria. It should be noted that the above criteria could not be fully applied for the multi-modal papers due to the small number of studies. Table 3 summarizes the studies that used only MRI features and Table 4 those that used multi-modal feature. Table 5 summarizes the rest of the studies which matched the original search criteria.

Studies based only on MRI features

The first study by Wee et al. [42] obtained an accuracy of 92.4% (sensitivity and specificity of 90.4% and 94.3%) for classification of AD and CTL subjects and an accuracy of 75.1% (sensitivity and specificity of 63.5% and 84.4%) for classification of converters and non-converters MCI subjects within 36 months. These results were achieved by using a multi-kernel SVM method and an integration of ROI-based and correlative morphological MRI features from 509 AD, MCI, and CTL subjects. The authors, in addition to ROI-based morphological features (including regional mean cortical thickness, cerebral cortical grey matter and cortical associated white matter volumes), proposed and used a new type of feature based on the measures of relative morphological abnormalities between different ROIs (named correlative features). Each correlative feature is a similarity of regional mean cortical thicknesses between a pair of ROIs. They also employed two types of feature selection methods to select the most discriminative features. Initially, two filter-based methods were used to reduce the number of features and subsequently, a wrapper-based approach (support vector machine recursive feature elimination (SVM-RFE) [97]) was used to select features that optimize the classifier performance. It has been shown that the integration of ROI-based and correlative morphological features resulted in a higher accuracy compared to each individual feature.

The second study obtained an accuracy of 91.5% (sensitivity and specificity of 89.8% and 95.6%) for AD classification and an accuracy of 75.9% (sensitivity and specificity of 64.0% and 66.9%) for MCI prediction [50]. In the study, a total of 259 features were extracted from MRI data of 699 AD, MCI, and CTL subjects and used as input to the OPLS classifier. Features included 34 cortical measures (7 type of measures from each region: cortical volume, cortical thickness, grey matter volume, surface area, mean curvature, Gaussian curvature, folding index and curvature index) and 21 regional volumes, which were automatically extracted with the FreeSurfer pipeline. Moreover, the authors investigated the effect of different normalization approaches on multivariate analysis performance. The best AD classification accuracy was obtained by combining raw cortical thickness measures with subcortical volumes normalized by intracranial volume, while using further features had no significant improvement. The authors did not perform feature selection, but excluded some measures which caused non-significant models.

Wolz et al. [17] utilized four different automated feature extraction techniques (namely hippocampal volume, TBM, cortical thickness, and manifold-based learning) to analyze structural MRI data of 834 ADNI AD, MCI, and CTL subjects. The extracted features were used to compare the performance of two classifiers, LDA and SVM, for AD classification and MCI prediction. The best accuracy for AD versus CTL classification was obtained by combining all extracted features and utilizing a LDA classifier; an accuracy of 89% (sensitivity and specificity of 93% and 85%). Similarly, using combined features and the LDA classifier resulted in the highest accuracy of 68% (sensitivity and specificity of 67% and 69%) for classification of MCI-converter and MCI-stable subjects. When different feature types were studied individually, the TBM features showed the best result. The authors also performed feature selection using a stepwise regression method. Moreover, age and gender correction using a linear regression model was applied to remove disease-related effects of age and gender on the classification.

The fourth study [36] investigated whether applying the local linear embedding (LLE) algorithm [98] to volumetric and cortical measures of brain MRI could

Table 3
Summary of the selected articles which used MRI features for classification

Methods and Material

Article	Classifier	Input Features	Feature selection	Validation	Number of subjects			
					AD	MCI-c	MCI-s	CTL
Wee et al. [42]	Multi kernel Support Vector Machines	Correlative and ROI-based morphological features	Hybrid feature selection method	10-fold CV	198	89	111	200
Westman et al. [50]	Orthogonal Projection to Latent Structures	Regional volume, cortical thickness, grey matter volume, surface area, mean curvature, gaussian curvature, folding index and curvature index measures	Not performed	7-fold CV	187	87	200	225
Liu et al. [36]	LDA, Support Vector Machines, and Logistic Regression with Elastic Net	Local linear embedded MRI features of regional brain volume and cortical thickness,	Not Performed	CV	86	97	93	137
Wolz et al. [17]	LDA and Support Vector Machines	Four MRI features: 1) hippocampal volume, 2) TBM, 3) cortical thickness and 4- a novel technique based on manifold learning	Stepwise regression feature selection	Leave-N-out/random sub-sampling CV	198	167	238	231
Chincarini et al. [28]	Support Vector Machines	Intensity and textural MRI-based features extracted from 9 VOIs and filtered with 18 filters	Random Forest classifier	20-fold CV	144	136	166	189

Results

Article	AD versus CTL			MCI-c versus MCI-s			MCI versus CTL		
	Accuracy	Sens/Spec	AUC	Accuracy	Sens/Spec	AUC	Accuracy	Sens/Spec	AUC
Wee et al. [42]	92.4	90/94	0.97	75.1	63.5/84.4	0.84	83.7	83.6/84.0	0.92
Westman et al. [50]	91.5	89.8/92.9	0.96	69.3	75.9/66.5	0.75	–	–	–
Liu et al. [36]	90	86/93	–	69	66/72	–	82*	81/82*	–
Wolz et al. [17]	89	93/85	–	68	67/69	–	84*	86/82*	–
Chincarini et al. [28]	–	89/94	0.97	–	72/65	0.74	–	89/80*	0.92*

AD, Alzheimer's disease; MCI-c, mild cognitive impairment subjects that later convert to AD; MCI-s, mild cognitive impairment subjects that remain stable; CTL, control subjects; ROI, region of interest; CV, cross validation; LDA, linear discriminant analysis; TBM, tensor based morphometry; VOI, volume of interest; AUC, area under ROC curve; AD versus CTL, classification of AD and control subjects; MCI versus CTL, classification of MCI and control subjects; MCI-c versus MCI-s, prediction of MCI subjects (conversion to AD in future); *MCI-c versus CTL, classification of MCI-c and control subjects.

Table 4
Summary of the selected articles which used multimodal features for classification

Methods and Material								
Article	Classifier	Input Features	Feature selection	Validation	Number of subjects			
					AD	MCI-c	MCI-s	CTL
Zhang et al. [12]	Multi-modal kernel- Support Vector Machines	Volumetric features from MRI and PET, and values of CSF measures	Simple feature selection based-on *t*-test	10-fold CV	51	43	56	52
Hinrichs et al. [46]	Multi-Kernel learning	Voxel-wise feature from MRI and PET, CSF measures, cognitive scores and APOE genotype	Implicit feature selection based-on *t*-test	10-fold CV	48		119*	66
Westman et al. [51]	Orthogonal Projection to Latent Structures	Regional subcortical volumes and cortical thickness measures from MRI, and values of CSF measures	No feature selection	7-fold CV	96	81	81	111
Wolz et al. [25]	Support Vector Machines	A unified biomarker, created from imaging (MRI) and non-imaging modalities (CSF biomarkers, APOE genotype and a risk factor associated with AD), using manifold learning	No feature selection	Leave-25%-out CV	103	89	112	116
Kohannim et al. [26]	Support Vector Machines	3 numerical summaries from MRI, one numerical summary from FDG-PET, CSF biomarkers, APOE genotype, age, gender, and body mass index	No feature selection	Leave-one-out CV	158		264*	213

Results									
Article	AD versus CTL			MCI-c versus MCI-s			MCI versus CTL		
	Accuracy	Sens/Spec	AUC	Accuracy	Sens/Spec	AUC	Accuracy	Sens/Spec	AUC
Zhang et al. [12]	93.2	93.0/93.3	0.98	–	–	–	76.4	81.8/66	0.81
Hinrichs et al. [46]	92.4	86.7/96.6	0.98	–	–	0.77	–	–	–
Westman et al. [51]	91.8	88.5/94.6	0.96	68.5	74.1/63.0	0.76	77.6	72.8/84.7	0.88
Wolz et al. [25]	88	85/91	–	69	68/70	–	87**	87/88**	–
Kohannim et al. [26]	90.7	–	0.92	–	–	–	75.8	–	0.77

AD, Alzheimer's disease; MCI-c, mild cognitive impairment subjects that later convert to AD; MCI-s, mild cognitive impairment subjects that remain stable; CTL, control subjects; MRI, magnetic resonance imaging; PET, positron emission tomography; CSF, cerebrospinal fluid; CV, cross validation; AUC, area under ROC curve; AD versus CTL, classification of AD and control subjects; MCI versus CTL, classification of MCI and control subjects; MCI-c versus MCI-s, prediction of MCI subjects (conversion to AD in future); *Total number of MCI subjects; **MCI-c versus CTL, classification of MCI-c and control subjects.

Table 5
List of articles reviewed

Article	Classifier	Validation	Input Features	Feature selection	Cohort	Study Population	Result	
							AD versus CTL	MCI-c versus MCI-s
Adaszewski et al. [34]	SVM	Train/Test, LOO CV	MRI: voxel-wise whole-brain grey matter	Yes	ADNI	106/203/137	-/-/-	-/64/69
Aguilar et al. [18]	ANN, DT, OPLS, SVM	10-fold CV	MRI: volumetric and cortical thickness measures; demographics; APOE genotype	Yes	ANM	116/119/110	88/86/90	74/81/68
Aksu et al. [35]	SVM	Train/Test	MRI: voxel intensities of volumetric density images	Yes	ADNI	120/300/180	89/59/98	-/-/-
Batmanghelich et al. [31]	SVM	Train/Test + CV	MRI: transformed low dimensional generative-discriminative features	Yes	ADNI	54/238/63	89/-/-	-/85/40
Chincarini et al. [28]	SVM	20-fold CV	MRI: VOI-based intensity and textural features	Yes	ADNI	144/302/189	-/89/94	-/72/65
Cho et al. [61]	LDA	Train/Test	MRI: spatial frequency representation of cortical thickness	No	ADNI	128/203/160	-/82/93	-/63/76
Chu et al. [22]	SVM	10-fold CV	MRI: voxel-based segmented grey matter	Yes	ADNI	131/261/188	84/-/-	67/-/-
Costafreda et al. [41]	SVM	Train/Test	MRI: hippocampal shape morphology	No	ANM	71/103/88	-/-/-	80/77/80
Coupe et al. [64]	LDA	CV	MRI: structure grading values (SNIPE) and volumes	No	ADNI	198/405/231	89/84/93	71/70/72
Coupe et al. [63]	QDA	Train/Test	MRI: structure grading values (SNIPE) and volumes	No	ADNI	60/-/60	90/88/92	-/-/-
Cui et al. [9]	SVM	Train/Test + CV	MRI: volumetric and thickness measures; CSF biomarkers; neuropsychological measures	Yes	ADNI	96/143/111	-/-/-	67/96/48
Cuingnet et al. [21]	LR, SVM	Train/Test + CV	MRI: 10 methods (voxel-based, cortical thickness, hippocampus)	Yes	ADNI	137/210/162	-/81/95	-/62/69
Davatzikos et al. [33]	SVM	CV	MRI: spatial pattern of abnormalities for recognition of early AD (SPARE-AD); CSF	Yes	ADNI	57/239/63	-/-/-	62/84/51
Duchesne et al. [29]	LDA, QDA, SVM	LOO CV	MRI: eigenspace based on intensity and deformation fields of VOIs	No	CSG-DF ICBM	75/-/75 -/-/149	92/-/-	-/-/-
Duchesne et al. [65]	LDA	LOO CV	MRI: eigenspace based on intensity and deformation fields of VOIs	No	CSG-DF	75/31/75	-/-/-	81/70/100
Escudero et al. [19]	ANN, DT, LR, SVM	10-fold CV	MRI: volumetric, surface and thickness measures; Demographics; APOE genotype	Yes	ADNI	122/222/180	89/-/-	-/-/-
Eskildsen et al. [66]	LDA	Train/Test, LOO CV	MRI: ROI-based discriminative patterns of cortical thickness measurements, age	Yes	ADNI	194/ /226	87/83/92	81/79/83
Ewers et al. [71]	LR	Train/Test	MRI: hippocampus volume and entorhinal cortex thickness; CSF biomarkers; neuropsychological tests	No	ADNI	81/130/101	94/96/95	74/83/67
Fan et al. [32]	SVM	LOO CV	MRI: regional volumetric maps	Yes	ADNI	56/88/66	94/-/-	-/-/-

Table 5
(*Continued*)

Article	Classifier	Validation	Input Features	Feature selection	Cohort	Study Population	Result	
							AD versus CTL	MCI-c versus MCI-s
Ferrarini et al. [20]	SVM	Train/Test + LOO CV	MRI: volumetric and morphological measures of Hippocampus	Yes	CSG-DF	50/30/50	90/88/92	80/80/80
Hamou et al.	DT	Train/Test	MRI: volumetric and cortical thickness; demographics; APOE genotype	No	ANM	120/122/112	-/-/-	-/-/-
Hinrichs et al. [70]	LPboosting	2-fold CV	MRI, PET: voxel-wise with incorporated spatial relationships	Yes	ADNI	89/-/94	84/84/82	-/-/-
Hinrichs et al. [45]	MKL	10-fold CV	MRI: voxel-wise features; PET	Yes	ADNI	77/-/82	81/79/82	-/-/-
Hinrichs et al. [46]	MKL	10-fold CV	MRI; FDG-PET; CSF biomarkers; APOE; Neuro-psychological data	Yes	ADNI	48/119/66	92/87/97	0.7667*
Kohannim et al. [26]	SVM	Train/Test + LOO CV	MRI: volumetric measures; PET: numerical summary; CSF biomarkers; APOE genotype; demographics	No	ADNI	158/264/213	91/-/-	-/-/-
Koikkalainen et al. [72]	Regression-based	CV	MRI: features from multi-template TBM	Yes	ADNI	188/369/215	86/81/91	72/77/71
Koikkalainen et al. [73]	Linear regression	LNO CV	MRI: FreeSurfer, volumes and VBM; neuropsychological tests; APOE genotype	Yes	ADNI	191/378/217	88/-/-	69/-/-
Liu et al. [69]	Ensemble method	10-fold CV	MRI: spatially normalized tissue density maps of T1-weighted images	Yes	ADNI	198/225/229	91/86/95	-/-/-
Liu et al. [36]	LDA, LR, SVM	LOO CV	MRI: LLE regional volume and cortical thickness	No	ADNI	86/190/137	90/86/93	69/66/72
Mattila et al. [39]	Bayes classifier, LR, SVM	10-fold CV	MRI:; CSF biomarkers; neuropsychological scores	No	ADNI	163/344/199	-/-/-	-/-/-
McEvoy et al. [60]	LDA	LOO CV	MRI: volumetric and thickness measures	Yes	ADNI	84/175/139	89/83/93	-/-/-
McEvoy et al. [62]	QDA	LOO CV	MRI: volumetric measures and longitudinal changes	No	ADNI	164/317/203	-/85/93	-/-/-
Nho et al. [23]	SVM	7-fold CV	MRI: grey matter density (VBM), volumetric and cortical thickness values	Yes	ADNI	182/389/226	91/85/95	72/78/68
Querbes et al. [68]	DT	10-fold CV	MRI: normalized thickness index; neuropsychological scores	No	ADNI	130/122/130	85/-/-	73/75/69
Simpson et al. [30]	Ensemble SVM	LOO CV	MRI VBM and deformation TBM	No	ADNI	149/-/162	86/-/-	-/-/-
Spulber et al. [49]	OPLS	7-fold CV	MRI: volumetric and cortical thickness measures	No	ADNI + ANM	295/434/335	88/86/90	68/70/67
Varol et al. [37]	Ensemble SVM	Train/Test	MRI: voxel-wise features	Yes	ADNI	116/-/148	88/86/89	-/-/-
Vemuri et al. [27]	SVM	Train/Test, 4-fold CV	MRI: GM, WM, and CSF tissue densities; demographics; APOE genotype	Yes	ADRC/ ADPR	190/-/190	89/86/92	-/-/-
Wee et al. [42]	SVM	10-fold CV	MRI: ROI-based and correlative morphological features	Yes	ADNI	198/200/200	92/90/94	75/64/84
Westman et al. [50]	OPLS	7-fold CV	MRI: volumetric, thickness, surface area, curvature measures	No	ADNI	187/287/225	92/90/93	70/76/67

Table 5
(*Continued*)

Article	Classifier	Validation	Input Features	Feature selection	Cohort	Study Population	Result	
							AD versus CTL	MCI-c versus MCI-s
Westman et al. [52]	OPLS	7-fold CV	MRI: automated volumetric and thickness measures, visual rating, manual measurement of hippocampus	No	ANM	75/101/81	83/77/88	-/74/70
Westman et al. [11]	OPLS	7-fold CV	MRI: volumetric and thickness measures	No	ADNI + ANM	295/444/335	87/87/87	-/71/60
Westman et al. [48]	OPLS	Train/Test + 7-fold CV	MRI: Automated regional segmentation and manual outlining of the hippocampus	No	ANM	117/122/112	-/90/94	-/73/-
Westman et al. [51]	OPLS	7-fold CV	MRI: volumetric and thickness measures, CSF biomarkers	No	ADNI	96/162/111	92/89/95	78/73/85
Wolz et al. [17]	LDA, SVM	LNO CV	MRI: Four different features	Yes	ADNI	198/405/231	89/93/95	-/-/-
Wolz et al. [25]	SVM	Leave-25%-out CV	A unified biomarker from MRI, CSF biomarkers, and APOE genotype	No	ADNI	103/201/116	88/85/91	69/68/70
Yang et al. [24]	SVM	Train/Test	MRI: ICA-based features	No	ADNI OASIS	202/410/236 100/-/416	81/82/80	-/-/-
Young et al. [38]	Gaussian processes, SVM	Train/Test	MRI: grey matter density map; PET: mean activity within region; CSF biomarkers; APOE genotype	No	ADNI	63/143/73	-/-/-	68/90/52
Zhang et al. [43]	SVM-based	10-fold CV	MRI: volumetric measures; PET: volumetric measures; CSF biomarkers; clinical scores	Yes	ADNI	45/91/50	93/-/-	74/69/74
Zhang et al. [12]	SVM-based	10-fold CV	MRI: volumetric measures; PET: volumetric measures; CSF biomarkers	Yes	ADNI	51/99/52	93/93/93	-/-/-

Results are represented as accuracy/sensitivity/specificity. *Area under ROC curve (AUC). AD, Alzheimer's disease; MCI-c, mild cognitive impairment subjects that later convert to AD; MCI-s, mild cognitive impairment subjects that remain stable; CTR, controls; SVM, support vector machines; OPLS, orthogonal projection to latent structures; MRI, magnetic resonance imaging; PET, positron emission tomography; LDA, linear discriminant analysis; MKL, multi kernel learning; DT, decision trees; QDA, quadratic discriminant analysis; ANN, artificial neuronal networks; LLE, local linear embedded; ICA, independent component analysis; LR, logistic regression; CV, cross validation; LOO CV, leave-one-out cross validation; LNO CV, Leave-n-out cross validation; ROI, region of interest; VOI, volume of interest; TBM, tensor based morphometry; VBM, voxel-based morphometry; CSF, cerebrospinal fluid; GM, grey matter; WM, white matter; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANM, AddNeuroMed Neuroimaging dataset; CSG-DF, IRCSS Centro San Giovanni di Dio Fatebenefratelli; OASIS, The Open Access Series of Imaging Studies; ADRC/ADPR, Mayo Clinic Alzheimer's Disease Research Center/Alzheimer's disease Patient Registry; ICBM, International Consortium for Brain Mapping. AD versus CTL, classification of AD and control subjects; MCI-c versus MCI-s, prediction of MCI subjects as AD or control. For studies that performed several experiments and provided more than one result, the highest accuracy is reported in this table.

improve the accuracy of classification and prediction of 413 ADNI AD, MCI, and CTL subjects. LLE is an unsupervised learning algorithm that transforms high dimensional data to low dimensional data by considering local symmetries and also global nonlinear structure of the data. The authors trained three different classifiers (SVM, LDA, and logistic regression with elastic nets) with the embedded features from AD and CTL subjects and used the trained classifiers to predict conversion from MCI to AD at baseline, according to follow-ups over 3 years. They showed that LLE features significantly improved the performance of both classification and prediction, compared to original features. The highest accuracy of AD versus CTL classification (90%) was achieved using LLE features through SVM and elastic nets regression classifiers. The highest accuracy of MCI prediction (69%) was obtained using the elastic nets classifier.

Finally, Chincarini et al. [28] proposed a fully automated technique to extract discriminative features based on selected pathology-specific volumes of interest in order to compute a classification index. They assessed the accuracy of the classification index on prediction of conversion from MCI to AD within a time-frame of 2 years. The volumes of interest included seven structures from the temporal lobe and two control volumes. The volumes of interest were filtered with 18 different filters, which resulted in a set of high dimensional MRI-based intensity and textural features. The authors used a Random Forest algorithm to reduce the dimensionality of the feature set and select the most relevant and important features for classification. Subsequently a SVM classifier was used to compute the classification index. The authors investigated the performance of the proposed algorithm on a population of 635 AD, MCI, and CTL subjects from the ADNI cohort. An accuracy of 92% was reported for classification of AD and CTL subjects, with an accuracy of 68% for MCI prediction.

Studies based on multi-modality features

CSF biomarkers have recently been used in addition to structural MRI data to classify AD subjects and to predict conversion from MCI to AD at multiple time points [51]. This study included 369 AD, MCI, and CTL subjects from the ADNI cohort. The regional subcortical volumes (23 measures) and cortical thickness measures (34 measures) that were extracted from structural MRI data and CSF biomarkers (3 measures) were used as input features to the OPLS classifier. The final model was built on a hierarchical fusion approach. The authors showed that combining features from structural MRI and CSF biomarkers resulted in a higher discriminant accuracy compared to the results of each modality alone. Classification of AD versus CTL resulted in an accuracy of 91.8% (sensitivity and specificity of 88.5% and 94.6%) and prediction of conversion from MCI to AD resulted in an accuracy of 68.5% (sensitivity and specificity of 74.1% and 63.0%), using both MRI and CSF features.

Zhang and colleagues [12] used a multiple-kernel SVM method to combine the biomarkers of three modalities (MRI, FDG-PET, and CSF) for classifying 202 AD, MCI, and CTL subjects, also from the ADNI cohort [19]. The volume of grey matter tissue of 93 ROIs from MRI and PET modalities in addition to the original value of three CSF measures were used as input features to the classifier. A multimodal data fusion and classification method (multi-kernel SVM classifier) was introduced and utilized to integrate the three modalities for classification task. The classifier included a kernel combination method that combines different kernel matrices into a single kernel matrix, and a linear SVM classifier. A simple feature selection based-on the *t*-test was performed to select the most discriminative features. Feature selection resulted in a higher classification accuracy compared to using all features. An accuracy of 93.2% (sensitivity and specificity of 93.0% and 93.3%) was achieved for classification of AD versus CTL, using all three modalities. The authors also investigated classification of MCI converters and MCI non-converters at 18 months follow-up, finding that 91.5% of MCI converters and 73.4% of MCI non-converters were correctly classified using a combination of the three modalities.

Multiple imaging modalities and clinical data including structural MRI, FDG-PET, CSF biomarkers, neuropsychological status exam scores, and APOE genotype data have also been used to discriminate AD and CTL subjects and to predict conversion from MCI to AD [46]. In this way, the authors applied an extension of kernel-SVM, namely MKL, to a population of 233 subjects from ADNI cohort. The main idea of the MKL classifier is to combine multiple kernel matrices into a single kernel matrix in order to create a superior classifier. The kernel matrices were created from voxel-wise features (extracted from imaging modalities), cognitive scores, CSF measures, and APOE genotype. An implicit feature selection step was also performed using a *t*-test approach. The classification of AD and CTL subjects resulted in an accuracy of 92.4% (sensitivity and specificity of 86.7% and 96.6%), using all imaging modalities, biological

measures, and cognitive scores). Based on the latter classifier, the multi-modality disease marker was assigned to individual MCI subjects.

Recently Wolz et al. [25] have proposed a framework based-on manifold learning to extract features from imaging modalities and combine them with non-imaging metadata. The result is a unified biomarker that can be used for data analysis and visualization. The authors investigated the performance of the proposed method on classification of 420 AD, MCI, and CTL subjects from the ADNI cohort. MRI was employed as the imaging modality and CSF biomarkers and APOE genotype were used as non-imaging metadata. A linear SVM classifier was utilized to distinguish different diagnostic groups. The classification of AD versus CTL subjects using all imaging and non-imaging data resulted in an accuracy of 88% (sensitivity and specificity of 85% and 91%), and the classification of MCI converters versus stable MCI subjects resulted in an accuracy of 69% (sensitivity and specificity of 68% and 70%).

Finally, Kohannim et al. [64] investigated the discriminative power of different biomarkers for classification of AD subjects and prediction of MCI conversion at one-year follow-up. A linear SVM classifier was employed to discriminate different groups of subjects using the following data: MRI, FDG-PET, CSF biomarkers, APOE genotype, age, gender, and body mass index. Numerical summary measures of hippocampal, ventricular, and temporal lobe volumes were used as MRI features. In addition, a numerical summary based on a predefined temporal lobe ROI was used as a FDG-PET feature. The total dataset included 635 AD, MCI, and CTL subjects from the ADNI cohort; however, three different subsets were defined according to data availability of each modality. The best AD versus CTL classification accuracy was 90.7%, which was obtained using MRI, FDG-PET, and CSF biomarkers.

DISCUSSION

Large amounts of data are generated from today's advanced image analysis pipelines. Image processing algorithms provide tools to extract relevant information from imaging data. Advances in multivariate data analysis and machine learning allow the combination of multiple variables from different modalities without having to deal with the problem of multiple comparisons. Patterns of disease can be observed rather than changes in single biomarkers. In recent years, many multivariate data analysis and machine learning studies have been published in the field of AD. Many different methods exists with SVM being the most popular followed by methods such as LDA, QDA, OPLS, ANN, and DT. As described above, the main problems when comparing different classifiers are that several factors including methods and cohort properties influence the accuracy. Analytical factors such as feature extraction methods, feature selection and the robustness of the validation approaches also affect the output. Further, dataset properties such as image quality, the number of subjects, demographics and clinical diagnosis criteria are important.

In the current paper, we have reviewed recent studies in the field of AD classification and MCI prediction that have used machine learning and multivariate data analysis. We have focused on studies that used structural MRI. Studies that used PET and CSF in addition to MRI have also been included, since these techniques are part of the new diagnostic criteria for AD. Additional studies have used other modalities such as functional MRI [99], only FDG-PET [100], diffusion tensor imaging [101, 102], MRI and magnetic resonance spectroscopy [103], MRI and vitamin E [104], MRI and functional MRI [105], as well as MRI and magnetoencephalography [106]. The latter studies, in addition to other studies that used a lower number of subjects than our limit of 100 subjects per study [107–111], or used classical statistical methods rather than multivariate or machine learning approaches [71, 112–114], or performed only MCI classification [115, 116], have not been discussed here.

In this article, we have presented key areas of multivariate analysis and machine learning including feature extraction, feature selection, classification, validation, and cohorts. As can be observed in Table 5, many different methods and algorithms have been proposed and employed. We have also discussed a selection of MRI and multimodal studies in detail (Tables 3 and 4). Even though it is difficult to compare the performance of the different classifiers due to use of different features and datasets, the classification accuracy for distinguishing between AD subjects and CTL individuals tends to be between 80–90%, and that for predicting conversion to MCI to AD somewhat lower. The accuracies are rarely higher for studies using large sample sizes and fully CV and/or external validation sets. Studies achieving close to 100% accuracies are likely to have an over-fitted model, result from very small homogenous samples or they may use clinical measures used for diagnosing the subjects as input, leading to circularity problems. With the same input, using the same cohort, different multivariate classifiers tend

to perform very similarly [18]. The results described above from the five ADNI studies utilizing different techniques further reinforce this. Very similar classification accuracies and predictions were obtained for the five studies. The small changes observed between the studies may be dependent on the slightly different numbers of subjects included. The lack of perfect performance may not lie with the method used for classification, but rather due to the clinical diagnosis not being 100% accurate. Neuropathology is by many considered to be the gold standard that can confirm the clinical diagnosis. There are studies that have utilized neuropathologically confirmed data as input for the multivariate classifier (SVM) [117]. These studies obtained very high accuracies although not as high as 100%. It has, however, been suggested that neuropathology should only be used as another biomarker, rather than the gold standard [118]. These studies also have very small sample sizes, due to the difficulty in obtaining neuropathologically-confirmed data and such small sample sizes can lead to over-fitting. Since many of the multivariate/machine learning studies of today are performed using the ADNI cohort, the performance of different classifiers are likely to be compared to many of these. It is important to consider that the ADNI cohort is a highly selected cohort with subjects that are very well educated. For this reason, studies based on other cohorts may obtain different results, which may be due to the choice of cohort rather than the methodology used [11]. This reinforces the need for results to be replicated using more than one cohort.

Standardization of the way research studies are performed is important to make studies more comparable. A recent article recommends that when the ADNI cohort is utilized the same subjects should be used for each study according to standardized lists that have been provided [86]. This study also recommends different ways of standardizing the analysis such as the use of 10-fold CV as mentioned above. Further, to standardize MRI protocols so data from different scanners can be combined as in the ADNI study would be very useful [119]. The AddNeuroMed study is harmonized to be comparable with ADNI [93] and can be combined and compared with ADNI as well as being used as an external validation set or for training the models [11]. Such an approach allows larger and more robust models as well as providing a good platform for validating the performance of different multivariate data analysis and machine learning techniques. Further studies have compared the performance of multivariate techniques with visual assessment by experienced radiologists [52, 117], which is a valuable approach for comparing such novel techniques to state of the art clinical diagnosis.

Considering the results of single-modality and multi-modality studies, the benefit of higher accuracy compared to the financial cost of acquiring additional biomarkers is relevant [17]. However, due to the complexity and heterogeneity of AD and other neurological disorders combining biomarkers reflecting different aspects of the disease are likely to be more effective for early diagnosis and predicting conversion from the prodromal stages of the disease. Particularly, in the case of prediction of conversion from MCI, which is of high importance and more difficult than classification of AD subjects, multimodality approaches may be worth using. It is however important to highlight that most studies using multimodal data obtain better results when combining the different modalities, but due to financial cost and availability of the different modalities, these studies include a significantly lower number of subjects than the largest studies only utilizing MRI. Comparing classification accuracies for AD versus CTL between multimodal studies and the largest MRI only studies, they are very similar. However, the ability to predict MCI conversion tends to be somewhat higher for the multimodal studies, which warrants further work to investigate the most optimal combination of biomarkers.

Several studies have also utilized different multivariate/machine learning techniques to create an index or score describing the pattern of disease, which can potentially be used and incorporated into clinical practice [33, 39, 49]. To find an easy way for the clinician to interpret diverse patient data in a concise and easy to understand manner is of great importance to improve AD diagnosis [39]. This type of approach could also be used to predict future MCI conversion [120–124] and aid in differential diagnosis [125].

FUTURE PERSPECTIVES AND CONCLUSIONS

A number of the machine learning or multivariate methods of today seem to perform with sufficient accuracy for both AD classification and MCI conversion prediction and the better performing of these provide broadly similar results. Limitations in accuracy may well lie with the data used for modeling (clinical diagnosis, input data, and cohort) rather than being a limitation of the methodology itself. We believe that models should be trained using all available AD versus CTL data (using CV) to build robust and representative models, which can subsequently be used to predict

MCI conversion. Some methods require feature selection methods when using high dimensional input, but we believe that it is more difficult to interpret results when using voxel based input, rather than ROI input for analysis. A limited number of discrete ROIs that are known to be affected early in AD may have greater face validity as a biomarker of disease than a complex pattern of voxels across the cortex, where the signal intensity of one voxel is indicative of AD but that of a neighboring voxel is not [62]. Data, which is easily interpreted, can be more easily used and implemented into clinical practice. Further, using ROIs as input, the numbers generated by any feature extraction method rarely reach levels of high dimensionality so feature selection should not be needed. We do not yet have a full understanding of the potential multiple patterns of AD-related atrophy, and it is possible that several subtypes of AD exist with similar clinical manifestation. Therefore we believe that a very limited set of predefined features may not be representative since the features may not be able to reflect the spatial-temporal pattern of structural and physiological abnormalities in their entirety [126]. It is important to further investigate the optimal combination of different biomarkers, which is of particular importance for predicting future MCI conversion to AD. Clinical measures need to be incorporated into models in an appropriate way to avoid circularity and over-fitting. Finally, confounding factors such as age, education, APOE, etc., need to be considered and incorporated where beneficial.

To conclude, different multivariate and machine learning techniques need to be carefully tested and validated against conventional diagnosis by experienced clinicians in a clinical setting and not only in highly selective research cohorts with strict inclusion and exclusion criteria. Finally, a simple way of describing the patterns of disease such as a disease index has the potential to be very useful. We believe multivariate analysis and machine learning have a great potential for being implemented in clinical practice to aid AD diagnosis, but also to target the right populations for clinical trials.

ACKNOWLEDGMENTS

Thanks to Swedish Brain Power, the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro). the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, The Swedish Medical Society, Loo och Hans Ostermans stiftelse för medicinsk forskning, Stiftelsen för ålderssjukdomar vid Karolinska Institutet, Karolinska Institutets forskningsbidrag and Axel och Signe Lagermans donationsstiftelse. Andrew Simmons was supported by funds from the NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London, the European Medical Information Framework and Alzheimer Research UK.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=2185).

REFERENCES

[1] Jack CR, Jr., Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH (2011) Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 257-262.

[2] Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353-356.

[3] Jack CR, Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-128.

[4] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Jr., Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (2011) Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 280-292.

[5] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.

[6] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[7] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[8] Carrillo MC, Dean RA, Nicolas F, Miller DS, Berman R, Khachaturian Z, Bain LJ, Schindler R, Knopman D (2013) Revisiting the framework of the National Institute

on Aging-Alzheimer's Association diagnostic criteria. *Alzheimers Dement* **9**, 594-601.

[9] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS, the Alzheimer's Disease Neuroimaging I (2011) Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors. *PLoS ONE* **6**, e21896.

[10] Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler DJ, Jr., Jennings RG, Karow D, Dale AM (2010) Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *AJNR Am J Neuroradiol* **31**, 347-354.

[11] Westman E, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Weiner MW, Lovestone S, Spenger C, Wahlund LO (2011) AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage* **58**, 818-828.

[12] 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.

[13] Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M (2010) Alzheimer's disease: Clinical trials and drug development. *Lancet Neurol* **9**, 702-716.

[14] Haller S, Lovblad KO, Giannakopoulos P (2011) Principles of classification analyses in mild cognitive impairment (MCI) and Alzheimer disease. *J Alzheimers Dis* **26**(Suppl 3), 389-394.

[15] Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (ACM, Pittsburgh, Pennsylvania, USA), pp. 144-152.

[16] Vapnik VN (1995) *The Nature of Statistical Learning Theory*, Springer, New York.

[17] Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, Soininen H, Lotjonen J, Alzheimer's Disease Neuroimaging I (2011) Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS ONE* **6**, e25446.

[18] Aguilar C, Westman E, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Simmons A, Wahlund LO (2013) Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Res* **212**, 89-98.

[19] Escudero J, Zajicek JP, Ifeachor E, Alzheimer's Disease Neuroimaging I (2011) Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials. *Conf Proc IEEE Eng Med Biol Soc* **2011**, 7957-7960.

[20] Ferrarini L, Frisoni GB, Pievani M, Reiber JH, Ganzola R, Milles J (2009) Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. *J Alzheimers Dis* **17**, 643-659.

[21] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, Chupin M, Benali H, Colliot O, Alzheimer's Disease Neuroimaging I (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.

[22] Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* **60**, 59-70.

[23] Nho K, Shen L, Kim S, Risacher SL, West JD, Foroud T, Jack CR, Weiner MW, Saykin AJ (2010) Automatic prediction of conversion from mild cognitive impairment to probable Alzheimer's disease using structural magnetic resonance imaging. *AMIA Annu Symp Proc* **2010**, 542-546.

[24] Yang W, Lui RL, Gao JH, Chan TF, Yau ST, Sperling RA, Huang X (2011) Independent component analysis-based classification of Alzheimer's disease MRI data. *J Alzheimers Dis* **24**, 775-783.

[25] Wolz R, Aljabar P, Hajnal JV, Lotjonen J, Rueckert D, Alzheimer's Disease Neuroimaging I (2012) Nonlinear dimensionality reduction combining MR imaging with non-imaging information. *Med Image Anal* **16**, 819-830.

[26] Kohannim O, Hua X, Hibar DP, Lee S, Chou Y-Y, Toga AW, Jack CR, Jr., Weiner MW, Thompson PM (2010) Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiol Aging* **31**, 1429-1442.

[27] Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR, Jr. (2008) Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *Neuroimage* **39**, 1186-1197.

[28] Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, Rei L, Squarcia S, Rodriguez G, Bellotti R, Cerello P, De Mitri I, Retico A, Nobili F, Alzheimer's Disease Neuroimaging I (2011) Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage* **58**, 469-480.

[29] Duchesne S, Caroli A, Geroldi C, Barillot C, Frisoni GB, Collins DL (2008) MRI-based automated computer classification of probable AD versus normal controls. *IEEE Trans Med Imaging* **27**, 509-520.

[30] Simpson IJ, Woolrich MW, Andersson JL, Groves AR, Schnabel JA (2013) Ensemble learning incorporating uncertain registration. *IEEE Trans Med Imaging* **32**, 748-756.

[31] Batmanghelich NK, Taskar B, Davatzikos C (2012) Generative-discriminative basis learning for medical imaging. *IEEE Trans Med Imaging* **31**, 51-69.

[32] Fan Y, Batmanghelich N, Clark CM, Davatzikos C (2008) Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* **39**, 1731-1743.

[33] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* **32** 2322, e2319-e2327.

[34] Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B, Initiative AsDN (2013) How early can we predict Alzheimer's disease using computational anatomy? *Neurobiol Aging* **34**, 2815-2826.

[35] Aksu Y, Miller DJ, Kesidis G, Bigler DC, Yang QX (2011) An MRI-derived definition of MCI-to-AD conversion for long-term, automatic prognosis of MCI patients. *PLoS ONE* **6**, e25074.

[36] Liu X, Tosun D, Weiner MW, Schuff N, Alzheimer's Disease Neuroimaging I (2013) Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. *Neuroimage* **83**, 148-157.

[37] Varol E, Gaonkar B, Erus G, Schultz R, Davatzikos C (2012) Feature ranking based nested support vector machine ensemble for medical image classification. *Proc IEEE Int Symp Biomed Imaging*, 146-149.

[38] Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S, Alzheimer's Disease Neuroimaging I (2013) Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin* **2**, 735-745.

[39] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van Gils M, Waldemar G, Soininen H, Lotjonen J, Alzheimer's Disease Neuroimaging I (2011) A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* **27**, 163-176.

[40] Vapnik VN (2012) *Estimation of Dependences Based on Empirical Data. Empirical Inference Science: Afterword of 2006*. New York.

[41] Costafreda SG, Dinov ID, Tu Z, Shi Y, Liu CY, Kloszewska I, Mecocci P, Soininen H, Tsolaki M, Vellas B, Wahlund LO, Spenger C, Toga AW, Lovestone S, Simmons A (2011) Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *Neuroimage* **56**, 212-219.

[42] Wee C-Y, Yap P-T, Shen D, for the Alzheimer's Disease Neuroimaging I (2013) Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Hum Brain Mapp* **34**, 3411-3425.

[43] Zhang D, Shen D, Alzheimer's Disease Neuroimaging I (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* **59**, 895-907.

[44] Lanckriet GRG, Cristianini N, Bartlett P, El Ghaoui L, Jordan MI (2004) Learning the kernel matrix with semidefinite programming. *J Mach Learn Res* **5**, 27-72.

[45] Hinrichs C, Singh V, Xu G, Johnson S (2009) MKL for robust Multi-modality AD Classification. *Med Image Comput Assist Interv* **5762**, 786-794.

[46] Hinrichs C, Singh V, Xu G, Johnson SC, Alzheimers Disease Neuroimaging I (2011) Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage* **55**, 574-589.

[47] Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). *J Chemom* **16**, 119-128.

[48] Westman E, Simmons A, Zhang Y, Muehlboeck JS, Tunnard C, Liu Y, Collins L, Evans A, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Wahlund LO (2011) Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage* **54**, 1178-1187.

[49] Spulber G, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Spenger C, Lovestone S, Wahlund LO, Westman E, dNeuroMed c, for the Alzheimer Disease Neuroimaging I (2013) An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment. *J Intern Med* **273**, 396-409.

[50] Westman E, Aguilar C, Muehlboeck JS, Simmons A (2013) Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. *Brain Topogr* **26**, 9-23.

[51] Westman E, Muehlboeck JS, 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.

[52] Westman E, Cavallin L, Muehlboeck JS, Zhang Y, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Spenger C, Lovestone S, Simmons A, Wahlund LO (2011) Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS ONE* **6**, e22506.

[53] Wold S, Ruhe A, Wold H, Dunn WJ (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J Sci Comput* **5**, 735-743.

[54] Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: A basic tool of chemometrics. *Chemometr Intell Lab Syst* **58**, 109-130.

[55] Trygg J, Wold S (2003) O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. *J Chemom* **17**, 53-64.

[56] Wold S, Trygg J, Berglund A, Antti H (2001) Some recent developments in PLS modeling. *Chemometr Intell Lab Syst* **58**, 131-150.

[57] Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J (2008) Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal Chem* **80**, 115-122.

[58] Friedman JH (1989) Regularized discriminant analysis. *J Am Stat Assoc* **84**, 165-175.

[59] Hastie T, Tibshirani R, Friedman JH (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY.

[60] McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ, Jr., Holland D, Karow DS, Pung CJ, Brewer JB, Dale AM, Alzheimer's Disease Neuroimaging I (2009) Alzheimer disease: Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology* **251**, 195-205.

[61] Cho Y, Seong JK, Jeong Y, Shin SY, Alzheimer's Disease Neuroimaging I (2012) Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* **59**, 2217-2230.

[62] McEvoy LK, Holland D, Hagler DJ, Fennema-Notestine C, Brewer JB, Dale AM (2011) Mild cognitive impairment: Baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology* **259**, 834-843.

[63] Coupe P, Eskildsen SF, Manjon JV, Fonov VS, Collins DL, Alzheimer's disease Neuroimaging I (2012) Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. *Neuroimage* **59**, 3736-3747.

[64] Coupe P, Eskildsen SF, Manjon JV, Fonov VS, Pruessner JC, Allard M, Collins DL, Alzheimer's Disease Neuroimaging I (2012) Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *Neuroimage (Amst)* **1**, 141-152.

[65] Duchesne S, Bocti C, De Sousa K, Frisoni GB, Chertkow H, Collins DL (2010) Amnestic MCI future clinical status prediction using baseline MRI features. *Neurobiol Aging* **31**, 1606-1617.

[66] Eskildsen SF, Coupe P, Garcia-Lorenzo D, Fonov V, Pruessner JC, Collins DL, Alzheimer's Disease Neuroimaging I (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* **65**, 511-521.

[67] Hamou A, Simmons A, Bauer M, Lewden B, Zhang Y, Wahlund L-O, Westman E, Pritchard M, Kloszewska I, Mecozzi P, Soininen H, Tsolaki M, Vellas B, Muehlboeck S, Evans A, Julin P, Sjögren N, Spenger C, Lovestone S, Gwadry-Sridhar F, Hamou A, Simmons A, Bauer M, Lewden B, Zhang Y, Wahlund L-O, Westman E, Pritchard M, Kloszewska I, Mecozzi P, Soininen H, Tsolaki M, Vellas B, Muehlboeck S, Evans A, Julin P, Sjögren N, Spenger

C, Lovestone S, Gwadry-Sridhar F (2011) Cluster analysis of MR imaging in Alzheimer's disease using decision tree refinement. *Int J Artif Intell* **6**, 90-99.

[68] Querbes O, Aubry F, Pariente J, Lotterie J-A, Démonet J-F, Duret V, Puel M, Berry I, Fort J-C, Celsis P, Initiative AsDN (2009) Early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve. *Brain* **132**, 2036-2047.

[69] Liu M, Zhang D, Shen D (2012) Ensemble sparse classification of Alzheimer's disease. *Neuroimage* **60**, 1106-1116.

[70] Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC, Alzheimer's Disease Neuroimaging I (2009) Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* **48**, 138-149.

[71] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, Jr., Feldman HH, Bokde AL, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging I (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.

[72] Koikkalainen J, Lotjonen J, Thurfjell L, Rueckert D, Waldemar G, Soininen H, Alzheimer's Disease Neuroimaging I (2011) Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *Neuroimage* **56**, 1134-1144.

[73] Koikkalainen J, Polonen H, Mattila J, van Gils M, Soininen H, Lotjonen J, Alzheimer's Disease Neuroimaging I (2012) Improved classification of Alzheimer's disease data via removal of nuisance variability. *PLoS ONE* **7**, e31112.

[74] Breiman L (1984) *Classification and Regression Trees*. Chapman & Hall.

[75] Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21-36.

[76] Ashburner J, Friston KJ (2000) Voxel-based morphometry–the methods. *Neuroimage* **11**, 805-821.

[77] Ashburner J, Hutton C, Frackowiak R, Johnsrude I, Price C, Friston K (1998) Identifying global anatomical differences: Deformation-based morphometry. *Hum Brain Mapp* **6**, 348-357.

[78] Chung MK, Worsley KJ, Paus T, Cherif C, Collins DL, Giedd JN, Rapoport JL, Evans AC (2001) A unified statistical approach to deformation-based morphometry. *Neuroimage* **14**, 595-606.

[79] Davatzikos C, Vaillant M, Resnick SM, Prince JL, Letovsky S, Bryan RN (1996) A computerized approach for morphological analysis of the corpus callosum. *J Comput Assist Tomogr* **20**, 88-97.

[80] Chupin M, Gerardin E, Cuingnet R, Boutet C, Lemieux L, Lehericy S, Benali H, Garnero L, Colliot O, Alzheimer's Disease Neuroimaging I (2009) Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* **19**, 579-587.

[81] Lotjonen J, Wolz R, Koikkalainen J, Julkunen V, Thurfjell L, Lundqvist R, Waldemar G, Soininen H, Rueckert D, Alzheimer's Disease Neuroimaging I (2011) Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. *Neuroimage* **56**, 185-196.

[82] Vallet B, Levy B (2008) Spectral geometry processing with manifold harmonics. *Computer Graphics Forum* **27**, 251-260.

[83] Shen D, Davatzikos C (2003) Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *Neuroimage* **18**, 28-41.

[84] Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR, Jr., Ashburner J, Frackowiak RSJ (2008) Automatic classification of MR scans in Alzheimer's disease. *Brain* **131**, 681-689.

[85] Schulerud H, Albregtsen F (2004) Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. *Comput Methods Programs Biomed* **73**, 91-99.

[86] Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, Cole PE, Crane PK, DeCarli C, Fox NC, Gunter JL, Hill D, Killiany RJ, Pachai C, Schwarz AJ, Schuff N, Senjem ML, Suhy J, Thompson PM, Weiner M, Jack CR, Jr., Alzheimer's Disease Neuroimaging I (2013) Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement* **9**, 332-337.

[87] Lehtovirta M, Soininen H, Laakso MP, Partanen K, Helisalmi S, Mannermaa A, Ryynänen M, Kuikka J, Hartikainen P, Riekkinen PJ (1996) SPECT and MRI analysis in Alzheimer's disease: Relation to apolipoprotein E epsilon 4 allele. *J Neurol Neurosurg Psychiatry* **60**, 644-649.

[88] Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* **2**, 189-210.

[89] Li L, Rakitsch B, Borgwardt K (2011) ccSVM: Correcting Support Vector Machines for confounding factors in biological data classification. *Bioinformatics* **27**, i342-i348.

[90] Dukart J, Schroeter ML, Mueller K (2011) Age correction in dementia–matching to a healthy brain. *PLoS ONE* **6**, e22193.

[91] Abdulkadir A, Mortamet B, Vemuri P, Jack CR, Jr., Krueger G, Kloppel S, Alzheimer's Disease Neuroimaging I (2011) Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. *Neuroimage* **58**, 785-792.

[92] 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) MRI measures of Alzheimer's disease and the AddNeuroMed Study. *Ann N Y Acad Sci* **1180**, 47-55.

[93] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Wahlund LO, Soininen H, Lovestone S, Evans A, Spenger C (2011) The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: Experience from the first 24 months. *Int J Geriatr Psychiatry* **26**, 75-82.

[94] Lovestone S, Francis P, Strandgaard K (2007) Biomarkers for disease modification trials–the innovative medicines initiative and AddNeuroMed. *J Nutr Health Aging* **11**, 359-361.

[95] Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund L-O, Ward M (2009) AddNeuroMed–The European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci* **1180**, 36-46.

[96] Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, Ivnik RJ, Tangalos EG, Petersen RC, Rocca WA (2008) The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology* **30**, 58-69.

[97] Rakotomamonjy A (2003) Variable selection using svm based criteria. *J Mach Learn Res* **3**, 1357-1370.

[98] Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323-2326.

[99] Tripoliti EE, Fotiadis DI, Argyropoulou M (2011) A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. *Artif Intell Med* **53**, 35-45.

[100] Gray KR, Wolz R, Heckemann RA, Aljabar P, Hammers A, Rueckert D, Alzheimer's Disease Neuroimaging I (2012) Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *Neuroimage* **60**, 221-229.

[101] Graña M, Termenon M, Savio A, Gonzalez-Pinto A, Echeveste J, Pérez JM, Besga A (2011) Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation. *Neurosci Lett* **502**, 225-229.

[102] Dyrba M, Ewers M, Wegrzyn M, Kilimann I, Plant C, Oswald A, Meindl T, Pievani M, Bokde ALW, Fellgiebel A, Filippi M, Hampel H, Klöppel S, Hauenstein K, Kirste T, Teipel SJ, group Es (2013) Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. *PLoS ONE* **8**, e64925.

[103] Westman E, Wahlund L-O, Foy C, Poppe M, Cooper A, Murphy D, Spenger C, Lovestone S, Simmons A (2010) Combining MRI and MRS to distinguish between Alzheimer's disease and healthy controls. *J Alzheimers Dis* **22**, 171-181.

[104] Mangialasche F, Westman E, Kivipelto M, Muehlboeck JS, Cecchetti R, Baglioni M, Tarducci R, Gobbi G, Floridi P, Soininen H, Kloszewska I, Tsolaki M, Vellas B, Spenger C, Lovestone S, Wahlund LO, Simmons A, Mecocci P (2013) Classification and prediction of clinical diagnosis of Alzheimer's disease based on MRI and plasma measures of alpha-/gamma-tocotrienols and gamma-tocopherol. *J Intern Med* **273**, 602-621.

[105] Abdulkadir A, Ronneberger O, Wolf RC, Pfleiderer B, Saft C, Kloppel S (2013) Functional and structural MRI biomarkers to detect pre-clinical neurodegeneration. *Curr Alzheimer Res* **10**, 125-134.

[106] Besga A, Ortiz L, Fernandez A, Maestu F, Arrazola J, Gil-Gregorio P, Fuentes M, Ortiz T (2010) Structural and functional patterns in healthy aging, mild cognitive impairment, and Alzheimer disease. *Alzheimer Dis Assoc Disord* **24**, 1-10.

[107] Dai Z, Yan C, Wang Z, Wang J, Xia M, Li K, He Y (2012) Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage* **59**, 2187-2195.

[108] Chaves R, Ramírez J, Górriz JM, López M, Salas-Gonzalez D, Álvarez I, Segovia F (2009) SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neurosci Lett* **461**, 293-297.

[109] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D, Initiative AsDN (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage* **65**, 167-175.

[110] Casanova R, Whitlow CT, Wagner B, Williamson J, Shumaker SA, Maldjian JA, Espeland MA (2011) High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Front Neuroinform* **5**, 22.

[111] Calvini P, Chincarini A, Gemme G, Penco MA, Squarcia S, Nobili F, Rodriguez G, Bellotti R, Catanzariti E, Cerello P, De Mitri I, Fantacci ME, Collaboration M, Initiative AsDN (2009) Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer disease. *Med Phys* **36**, 3737-3747.

[112] Schmand B, Eikelenboom P, van Gool WA, Alzheimer's Disease Neuroimaging I (2012) Value of diagnostic tests to predict conversion to Alzheimer's disease in young and old patients with amnestic mild cognitive impairment. *J Alzheimers Dis* **29**, 641-648.

[113] Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hanninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala EL, Vainio P, Vanninen R, Partanen K, Soininen H (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol Aging* **25**, 303-310.

[114] Grydeland H, Westlye LT, Walhovd KB, Fjell AM (201d) Improved prediction of Alzheimer's disease with longitudinal white matter/gray matter contrast changes. *Hum Brain Mapp* **34**, 2775-2785.

[115] Cui Y, Wen W, Lipnicki DM, Beg MF, Jin JS, Luo S, Zhu W, Kochan NA, Reppermund S, Zhuang L, Raamana PR, Liu T, Trollor JN, Wang L, Brodaty H, Sachdev PS (2012) Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: A combined spatial atrophy and white matter alteration approach. *Neuroimage* **59**, 1209-1217.

[116] Zhou L, Wang Y, Li Y, Yap PT, Shen D, Alzheimer's Disease Neuroimaging I (2011) Hierarchical anatomical brain networks for MCI prediction: Revisiting volumetric measures. *PLoS ONE* **6**, e21935.

[117] Kloppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, Mader I, Mitchell LA, Patel AC, Roberts CC, Fox NC, Jack CR, Jr., Ashburner J, Frackowiak RS (2008) Accuracy of dementia diagnosis: A direct comparison between radiologists and a computerized method. *Brain* **131**, 2969-2974.

[118] Scheltens P, Rockwood K (2011) How golden is the gold standard of neuropathology in dementia? *Alzheimers Dement* **7**, 486-489.

[119] Jack CR, Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, J LW, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**, 685-691.

[120] 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.

[121] Mattila J, Soininen H, Koikkalainen J, Rueckert D, Wolz R, Waldemar G, Lötjönen J (2012) Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. *J Alzheimers Dis* **32**, 969-979.

[122] Simonsen AH, Mattila J, Hejl A-M, Frederiksen KS, Herukka S-K, Hallikainen M, van Gils M, Lötjönen J, Soininen H, Waldemar G (2012) Application of the PredictAD software tool to predict progression in patients with mild cognitive impairment. *Dement Geriatr Cogn Disord* **34**, 344-350.

[123] Liu Y, Mattila J, Ruiz MÁM, Paajanen T, Koikkalainen J, van Gils M, Herukka S-K, Waldemar G, Lötjönen J, Soininen H, Initiative AsDN (2013) Predicting AD conversion: Comparison between prodromal AD guidelines and computer assisted PredictAD tool. *PLoS ONE* **8**, e55246.

[124] Muñoz-Ruiz MÁ, Hall A, Mattila J, Koikkalainen J, Herukka S-K, Vanninen R, Liu Y, Lötjönen J, Soininen H (2014) Comparing predictors of conversion to Alzheimer's disease using the disease state index. *Neurodegener Dis* **13**, 200-202.

[125] Muñoz-Ruiz MÁ, Hartikainen P, Hall A, Mattila J, Koikkalainen J, Herukka S-K, Julkunen V, Vanninen R, Liu Y, Lötjönen J, Soininen H (2013) Disease state fingerprint in frontotemporal degeneration with reference to Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis* **35**, 727-739.

[126] Fan Y, Resnick SM, Wu X, Davatzikos C (2008) Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. *Neuroimage* **41**, 277-285.

[127] Damangir S, Manzouri A, Oppedal K, Carlsson S, Firbank MJ, Sonnesyn H, Tysnes OB, O'Brien JT, Beyer MK, Westman E, Aarsland D, Wahlund LO, Spulber G (2012) Multispectral MRI segmentation of age related white matter changes using a cascade of support vector machines. *J Neurol Sci* **322**, 211-216.