HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 July 15.

Published in final edited form as:

Neuroimage. 2017 July 15; 155: 530–548. doi:10.1016/j.neuroimage.2017.03.057.

A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages

Saima Rathore[1], **Mohamad Habes**[1], **Muhammad Aksam Iftikhar**[2], **Amanda Shacklett**[1], and **Christos Davatzikos**[1,*]

[1]Center for Biomedical Image Computing and Analytics, Perelman School of Medicine, University of Pennsylvania, USA

[2]Department of Computer Science, Comsats Institute of Information technology, Lahore, Pakistan

Abstract

Neuroimaging has made it possible to measure pathological brain changes associated with Alzheimer's disease (AD) *in vivo*. Over the past decade, these measures have been increasingly integrated into imaging signatures of AD by means of classification frameworks, offering promising tools for individualized diagnosis and prognosis. We reviewed neuroimaging-based studies for AD classification and mild cognitive impairment, selected after online database searches in Google Scholar and PubMed (January, 1985 to June, 2016). We categorized these studies based on the following neuroimaging modalities (and sub-categorized based on features extracted as a post-processing step from these modalities): i) structural magnetic resonance imaging [MRI] (tissue density, cortical surface, and hippocampal measurements), ii) functional MRI (functional coherence of different brain regions, and the strength of the functional connectivity), iii) diffusion tensor imaging (patterns along the white matter fibers), iv) fluorodeoxyglucose positron emission tomography (metabolic rate of cerebral glucose), and v) amyloid-PET (amyloid burden). The studies reviewed indicate that the classification frameworks formulated on the basis of these features show promise for individualized diagnosis and prediction of clinical progression. Finally, we provided a detailed account of AD classification challenges and address some future research directions.

Keywords

Alzheimer's disease; Mild cognitive impairment; Machine learning; Classification; Neuroimaging; Feature extraction

[*]Correspondence to: Center for Biomedical Image Computing and Analytics, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, PA 19104, USA. Tel.: +215 746 4067. christos.davatzikos@uphs.upenn.edu.

1. Introduction

Alzheimer's disease (AD), the most prevalent form of dementia, is expected to affect 1 out of 85 people in the world by the year 2050 (Brookmeyer et al., 2007). The pathophysiology of AD is increasingly becoming clearer. The brain of an AD patient accumulates abnormal proteins (Aβ and tau) in the form of amyloid *plaques* and neurofibrillary tangles, eventually resulting in loss of neurons (Frisoni et al., 2010; Jagust, 2013). Brain changes due to AD occur even before amnestic symptoms appear (Buckner, 2004), and occur in a pattern that typically includes the temporal lobe and hippocampus (Braak, Braak, 1991). It has been suggested that this inevitable atrophy can be a valuable marker of neurodegeneration (Frisoni et al., 2010), as measured with structural magnetic resonance imaging (sMRI). Further alterations in function, connectivity and metabolism can be detected using functional MRI (fMRI) (Agosta et al., 2012; Binnewijzend et al., 2012; Dennis, Thompson, 2014; Fan et al., 2011; Fox, Raichle, 2007), and fluorodeoxyglucose positron-emission tomography (FDG-PET) (Gray et al., 2012; Padilla et al., 2012; Pagani et al., 2015; Teipel et al., 2015; Toussaint et al., 2012). However, the subtleties of the changes in early AD stages make it difficult to distinguish patterns easily by conventional radiologic readings or even by quantitative analysis. Thus, it remains challenging to establish reliable markers for diagnosing and monitoring disease progression in the early stages and on an individual basis.

Numerous neuroimaging studies have used region of interest (ROI)-types of analyses to investigate subtle changes associated with AD (Chetelat, Baron, 2003; Lerch et al., 2008). Such studies rely solely on prior knowledge to guide the selection of ROIs and features, thus ignoring brain changes outside the studied region(s) and failing to discover new knowledge. Machine learning offers a systematic approach in developing sophisticated, automatic, and objective classification frameworks for analyzing high-dimensional data and can learn complex and subtle patterns of change across various imaging modalities (Sajda, 2006). Typically, a classification framework includes at-least feature extraction and classification algorithm to build predictive models that facilitate the automation of medical decision support (Chiang, Pao, 2016) and provide increased objectivity in these decisions. Furthermore, classification frameworks can be used to develop imaging markers or indices (Davatzikos et al., 2008) with high sensitivity and specificity in individuals (Sajda, 2006) that can summarize the imaging profile of a subject into a single meaningful value (Habes et al., 2016b). This creates a more individualized, patient-tailored approach (Ithapu et al., 2015), which is imperative in the current age of personalized medicine because it allows further consideration of genetic or life-style risks, by utilizing advanced computational power (Habes et al., 2016a; Habes et al., 2016b; Habes et al., 2016c).

In recent years, a large body of research has been published on neuroimaging-based computer-aided classification of AD and its prodromal stage, mild cognitive impairment (MCI). Motivated by this rapid proliferation of AD/MCI classification studies and the lack of literature summarizing different AD-related features as extracted from neuroimaging data and classification algorithms, we present an overview of pertinent advances in this field. We summarize key representative studies on neuroimaging-based classification of AD/MCI and provide a brief account of the main aspects of these studies, such as study population, type of features, the adopted classification algorithm, and the reported classification success rates.

Furthermore, we highlight several bottlenecks (i.e. limited sample size and variability in data settings across the different studies) and discuss the generalizability and reproducibility of existing AD classification studies, as well as the important and largely unexplored issue of heterogeneity in AD.

Recent review papers (Arbabshirani et al., 2017; Falahati et al., 2014) reported studies on MRI-based classification of AD and MCI, limiting AD classification to MRI only. Pathological brain changes related to AD can be captured via various imaging modalities, such as FDG-PET and amyloid-PET, therefore, a comprehensive review on AD classification should not be limited to MRI only. This review is further unique in that it focuses exclusively on those studies that have extensively leveraged cross-validation strategies to estimate the performance of their classification frameworks. Cross-validation is generally designed to achieve independent training and test data for a classification algorithm and defined as split the data once (split-in-train-test) or several times (k-fold cross-validation) to obtain an unbiased estimate of the classification performance of the algorithm and avoid over fitting (Arlot, Celisse, 2010; Kohavi, 1995). In the split-in-train-test, data is randomly divided into independent training and test subsets, optimally with matched demographic characteristics. The training subset is used solely for the learning procedure of the classification algorithm and the test subset is used to estimate the performance of the trained classification algorithm. In k-fold, data is divided into k-folds and a classification algorithm is tested on k^{th} fold after being trained on k-1 folds in k^{th} iteration. Furthermore, we provide in-depth detail about AD-related feature extraction methods from various neuroimaging modalities, important information that is mostly lacking in existing review papers.

2. Selection criteria

We searched in PubMed and Google Scholar, from January 1985 to June 2016, and identified 409 studies based on the given search criteria. We included original peer-reviewed research studies that exclusively used cross-validation strategies to estimate the performance of their classification frameworks. In addition, studies conducted for method comparisons and studies not focusing primarily on AD classification were excluded from this review. Finally, this criterion resulted in 81 studies that were reviewed and presented here. A more thorough explanation of the search and screening process, and databases generated from the search in Google Scholar and PubMed are provided in the Supplementary Material.

3. Classification frameworks for Alzheimer's disease and its prodromal stages

Over the past decade, classification frameworks have been used successfully to analyze complex patterns in neuroimaging data with a view to the classification of AD and MCI subjects. A classification framework is comprised of four major components: feature extraction, feature selection, dimensionality reduction, and feature-based classification algorithm. Feature extraction and classification algorithm are the minimally required components, as shown in Figure 1, whereas other components can be applied as needed. The studies having minimal required components of classification framework were considered

potential candidates for inclusion in the paper provided meeting other criteria. In the feature-extraction process, AD-related features from various neuroimaging modalities, such as structural MRI, functional MRI, diffusion tensor imaging (DTI), amyloid-PET, and FDG-PET, are extracted from the training subjects. The term 'features' refers to the post-processing applied on raw medical imaging data to derive more informative measures. The examples of such derived measures include regional tissue densities, regional cortical thickness, etc. These derived measures can vary from millions (when all the voxels are used as features) to a few (when a few representative measures are extracted from the brain). The features extracted from various modalities can be used in isolation or combined to make use of the complementary information provided by several modalities. A classification algorithm (predictive model) is then trained on the extracted features to provide diagnostic support in predicting cognitively normal (CN) and diseased subjects.

In the aforementioned classification framework, the selection of an appropriate modality for obtaining imaging data and accurate feature extraction for AD classification is often more important than the selection of the underlying classification algorithms in order to achieve higher prediction rates (Sabuncu, Konukoglu, 2015). Therefore, we provided more details on feature extraction for the included studies in this review. Overall, the paper is divided into various sections, where each section focuses on features extracted from one particular imaging modality, such as structural MRI, functional MRI, DTI, and PET. A section on multimodal AD/MCI classification studies, which describes how features extracted from various imaging modalities are combined to utilize their complementary information, is included at the end.

3.1. Structural MRI-based studies

Cerebral neurodegeneration is characterized by early damage to synapses, followed by degeneration of axons and ultimately, atrophy of the dendritic tree and perikaryon (Serrano-Pozo et al., 1101). This neurodegeneration process is more severe in certain parts of the brain, such as the right and left hippocampus, temporal and cingulate gyri, and precuneus (Baron et al., 2001; Busatto et al., 2003; Frisoni et al., 2002; Ishii et al., 2005). The inevitable atrophy, caused by neurodegeneration, is generally measured using structural MRI, and serves as a valuable marker of the stage and aggressiveness of the neurodegenerative aspect of the AD pathology in the individual (Frisoni et al., 2010; Vemuri, Jack, 2010). The atrophic process in these regions leads to profound structural changes in the brain, such as thinning of the cortical surface, structural variation in several brain regions and variation in the regional tissue densities, and have been demonstrated in several neuroimaging-based studies of AD classification. A top-level breakdown of these studies is shown in Figure 2. Three main feature extraction methods for assessing structural variation are considered: i) density maps, ii) cortical surface, and iii) pre-defined regions-based methods.

3.1.1. Density map-based methods—Density map-based methods quantify patterns of atrophy by utilizing density maps of white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF), which are generated by methods such as voxel-based morphometry (VBM) (Ashburner, Friston, 2000) or regional analysis of volumes examined

in normalized space (RAVENS) maps (Davatzikos et al., 2001). Depending on how the density maps are used, these methods can be further divided into two categories: i) whole density maps as features (DMAF), and ii) reduced density map features. The studies that fall within these categories are listed in Table 1.

3.1.1.1. Whole density maps as features (DMAF)-based methods: This method is centered on the construction of a feature vector by utilizing the density maps for WM, GM, or both for classification.

In an earlier study, the GM density map of the entire brain together with a support vector machine (SVM) achieved promising AD classification (Kloppel et al., 2008). In that study, relatively lower GM density was found in the hippocampus of AD subjects, which was a strong indicator of hippocampal atrophy, consistent with previous research (Frisoni et al., 2002). In addition, GM maps have been used for AD classification, by employing a large-scale regularization approach (Casanova et al., 2011) and spatially augmented linear programming boosting method (LPBM) (Hinrichs et al., 2009), and to predict conversion from MCI to AD, by using SVM (Adaszewski et al., 2013). Termenon et al., used GM density maps to develop feature vectors for AD classification (Termenon, Graña, 2012) by employing a two-stage classification framework, wherein a relevance vector machine classifier (Tipping, 2001) was used in the first stage. The subjects that fell into the low confidence interval of the classifier were used as the input for the second classifier in the prediction. SVM, nearest neighbor, relevance vector machine, and learning vector quantization were used as second-stage classifiers, however, SVM was better. Recently, Moller et al. also used GM density map for SVM-based AD classification (Moller et al., 2016).

Furthermore, the Jacobian determinants, calculated from these density maps, have been used as features for predicting conversion from MCI to AD by using SVM, Bayes statistics, and voting feature interval classifiers (Plant et al., 2010).

These methods achieved a ≥ 81% accuracy in AD classification, and a ≥ 62% accuracy in prediction of AD conversion. Furthermore, the two-stage framework proposed by Termenon et al. demonstrated superior classification accuracy (from 77% to 87%) than its single-stage-based counterpart (Termenon, Graña, 2012).

3.1.1.2. Reduced density map feature-based methods: DMAF-based methods suffer the drawback of dimensionality, as the numbers of features are typically larger than, or comparable to, the number of the available subjects. When the number of features is high relative to the number of subjects in the training set, it is possible that classification rules yielding high accuracy on the training set were originated only by chance. This can lead many classification algorithms to select classification rules that could fail to generalize to new data (Vapnik, 1999). Consequently, features have been reduced using supervised or unsupervised feature-reduction methods, or they have been extracted from pre-defined atlases and adaptive regions in order to reduce dimensionality.

• ***Supervised/unsupervised feature-reduction-based methods:*** These methods essentially focus on distilling large-sized density maps to fewer, meaningful features in a supervised or unsupervised fashion. Among the unsupervised methods, Salvatore et al. used principal component analysis (PCA) to reduce the dimensions of WM and GM density maps. The reduced density maps were used for SVM-based AD classification and prediction of conversion from MCI to AD (Salvatore et al., 2015). In addition, Liu et al. used local linear embedding method to transform multivariate regional brain volume and cortical thickness MRI data to a locally linear space, with fewer dimensions, while also utilizing the global nonlinear data structure (Liu et al., 2013). The embedded brain features were then used to train classification algorithms such as regularized logistic regression (RLR), SVM, and linear discriminant analysis (LDA).

On the other hand, Beheshti et al. proposed reduction of the dimensions of GM maps in a supervised fashion (Beheshti, Demirel, 2015). They used the intensity distribution of voxels of GM maps, rather than using the intensities of all the voxels of GM maps, as features. The optimal number of bins in the intensity distribution was selected based on the Fisher criterion maximization between AD and CN subjects, and the resultant intensity distribution-based features were used for SVM-based AD classification.

• ***Atlas-based methods:*** These methods rely on parcellation of brain image into several anatomical regions based on pre-defined anatomically labeled atlases (such as automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and laboratory of neuroimaging (LONI) atlas (Shattuck et al., 2008)), followed by extraction of features from those particular regions. For example, Magnin et al., used AAL to parcellate the brain image into 116 regions, and then used the relative weight of the GM, compared to that of the WM and CSF, for each parcellated region to develop a feature vector for SVM-based AD classification (Magnin et al., 2009).

• ***Adaptive-ROI-based methods:*** Traditionally, atlas-based methods are used to obtain regional measurement of anatomical features and to investigate abnormal tissue structures in disease conditions. However, in practice, prior knowledge of abnormal regions is not always available. Even when a prior hypothesis can be made about specific ROIs, a region demonstrating abnormality might be part of a single ROI, or span multiple ROIs, thereby potentially reducing the significance of further analysis. Therefore, adaptive ROIs have been calculated to reduce the dimensions of density maps and to resolve this issue. Furthermore, depending on the number of sets of adaptive ROIs that are calculated, these methods can be divided into two categories: i) single-set adaptive ROIs, and ii) multiple-set adaptive ROIs.

1. Single-set adaptive ROIs: In this method, subjects are registered to one particular atlas, and adaptive ROIs and corresponding regional volumetric measures are calculated in that atlas space. In earlier work, Davatzikos et al. calculated RAVENS density maps and used a watershed clustering algorithm-based method (Fan et al., 2007) to calculate an adaptive set of ROIs for SVM-based AD classification, using the multi-centric ADNI (Alzheimer's disease neuroimaging initiative (Weiner et al., 2015)) dataset. They reported a high cross-validated classification accuracy of 94.30%, with a pattern involving many temporal lobe GM regions, peri-hippocampal WM, and CSF (Fan et al., 2008a). The trained SVM in this

study was used to determine the spatial pattern of abnormality for recognition of early AD (SPARE-AD) index, which was later tested independently in CN and MCI subjects of the Baltimore longitudinal study of aging (BLSA) (Davatzikos et al., 2011) dataset. In subsequent studies, the same classification framework was used for MCI classification (Davatzikos et al., 2008) and prediction of conversion from MCI to AD (Misra et al., 2009).

2. Multiple-set adaptive ROIs: In this method, subjects are registered to multiple atlases, and adaptive ROIs and corresponding regional volumetric measures are calculated in each atlas space to overcome the inherent bias associated with one atlas. For example, Min et al. derived multiple atlases from the non-overlapping clusters of subjects (Min et al., 2014), obtained using affinity propagation (Frey, Dueck, 2007). They registered subjects to the atlases and adaptively calculated a set of ROIs and volumetric features in each atlas space. The top-most K discriminating features calculated from each atlas space were combined for SVM-based classification. Subsequently, Liu et al. argued that the features extracted from K sets of adaptive ROIs are different representations of the same subject (Liu et al., 2015), and should not be concatenated, as in a previous study (Min et al., 2014). To resolve this, Liu et al. registered subjects to different selected atlases and extracted features from adaptive regions of each atlas-registered image, viewing that image as the main source, and all other atlas registered-images as adjunctive sources (Liu et al., 2015). SVM was separately trained on features extracted from each set and the results of multiple sets were combined using majority voting.

The multiple-set adaptive ROIs-based methods were quite effective, and improved AD/CN classification from 84.18% to 92.51% (Liu et al., 2015) and progressive MCI (pMCI)/stable MCI (sMCI) classification from 70.06% to 78.88% (Liu et al., 2015) compared to single-set adaptive ROIs-based methods.

3.1.2. Surface-based methods—AD patients generally show changes in temporal and parietal regions of the cortical surface (Bakkour et al., 2009; Dickerson et al., 2009; Dickerson et al., 2011). Although these changes are not easily visible or measureable in the early stages of AD, classification frameworks have been able to detect subtle changes in the cortical surface by analysis of complex cortical surface data, in a way that is complementary to regional volumetric maps extracted in a voxel-wise manner. Cortical surface measures are extracted from all the vertices of a surface. These measures are either used directly or are reduced by applying feature reduction methods, thereby leading to two main categories: i) vertices as features-based methods, and ii) reduced vertices as features-based methods. The studies of AD diagnosis support employing cortical surface-based features are listed in Table 2.

3.1.2.1. Vertices as features-based methods: This family of AD classification frameworks solely relies on the features calculated from all the vertices of a cortical surface. For instance, Li et al. calculated a variety of morphological features, including volumetric (cortical thickness, surface area, and GM volume) and geometric (sulcal depth, metric distortion, and mean curvature) measures, at each vertex on the pial surface (Li et al., 2014b), which were used for SVM-based MCI classification.

3.1.2.2. Reduced vertices as features-based methods: The use of cortical surface features of all the vertices suffers from the same dimensionality drawback, for reasons similar to those mentioned in Section 3.1.1.2. Consequently, these features have either been reduced by means of feature-reduction methods, or extracted from regions of pre-defined atlases.

• ***Supervised/unsupervised feature-reduction-based methods:*** In these methods, the dimensions of long feature vector comprising features of all the vertices of the brain are reduced by applying supervised or unsupervised feature reduction methods. For example, Cho et al. converted thickness data to a frequency domain and achieved lower dimensionality by filtering out high-frequency (noise) components (Cho et al., 2012). They employed an incremental learning-based LDA for reduced dataset-based AD classification. Park et al. addressed this issue by modeling the cortical surface using three-dimensional (3D) meshes and extracted cortical thickness parameterized by these meshes (Park et al., 2012) for SVM-based AD and MCI classification. Later, Park et al. applied this classification framework to longitudinal data for the early detection of AD (Park et al., 2013). They trained SVM on MCI and CN subjects, and tested it on the subjects who converted to AD, using the images of the subjects taken one time-point before the actual conversion. They achieved promising early prediction of conversion (83%) from MCI to AD.

• ***Atlas-based methods:*** In these methods, the original brain images are registered to certain standardized stereotaxic spaces (Fischl et al., 1999), and cortical maps/features are computed (Fischl, Dale, 2000; Jones et al., 2000; MacDonald et al., 2000) and tessellated into various regions using existing atlas templates (Desikan et al., 2006). Unlike the structural templates discussed in Section 3.1.1.2, which are used in the volume space, these atlas templates are used in the surface space. The features of the tessellated regions are used in feature vector development, which then is used in classification.

Desikan et al. used average cortical thickness of all the tessellated regions for MCI and AD classification (Desikan et al., 2009) using logistic regression (LR). The cortical thickness in the entorhinal cortex and supramarginal gyrus proved to be a better predictor of MCI than the cortical thickness of other regions. Similarly, Oliveira et al. used regional thickness measures and the average thickness of the entire brain for AD classification by means of SVM (Oliveira et al., 2010). They found that the average cortical thickness of the entire brain was a better predictor of AD than the regional thickness features. Wee et al. used regional thickness measures, cerebral cortical GM and associated WM volumes, and correlative features, which were obtained based on the similarity of cortical thickness between pairs of brain regions (Wee et al., 2013). Features were first selected using t-tests, and later using minimum-redundancy and maximum-relevance (mRMR) (Peng et al., 2005) in conjunction with SVM recursive-feature elimination (SVM-RFE). Multi-kernel SVM was used for classification. McEvoy et al. also used average regional cortical thickness and volumetric measures for LDA-based AD classification and prediction of conversion from MCI to AD (McEvoy et al., 2009). Lillemark et al. used the proximity between the center of mass and percentage surface connectivity of different brain regions as features for LDA-based AD and MCI classification (Lillemark et al., 2014).

Recently, Eskildson et al. investigated the prediction of AD conversion by measuring regional thickness at various stages of MCI conversion to AD (Eskildsen et al., 2013). To this end, pMCI subjects were categorized based on time to conversion to AD (6, 12, 24, or 36 months), and each category was classified against sMCI subjects. Features were reduced using mRMR, and LDA was used to determine the classification accuracy. The classification based on stage-specific categories of pMCI subjects demonstrated better accuracy than the overall classification of pMCI and sMCI subjects.

In addition, the efficacy of regional thickness measures was investigated using orthogonal partial least-squares of the latent structures for AD classification in two independent cohorts (ADNI and AddNeuroMed (Lovestone et al., 2009; Lovestone et al., 2007)) and a pool of these cohorts (ADNI + AddNeuroMed) (Westman et al., 2011). Results demonstrated similar patterns of atrophy in two individual cohorts, showing that the key regions involved in AD classification were similar in both cohorts. They further proved similarity in the patterns of atrophy through training the classification algorithm on one cohort and testing it on the other. For example, AD classification derived from training on the AddNeuroMed cohort and testing on the ADNI cohort, and vice versa, led to accuracies of 86.0% and 83.40%, respectively. The individual and pooled cohorts were further used to predict the conversion from MCI to AD. For instance, the classification algorithm trained on the combined cohort classified 71% of the pMCI as AD-like and 60% of the sMCI as CN-like.

The reduced vertices as features-based methods demonstrated better classification accuracy than raw vertices as feature-based methods as shown by improvement of 10–13% for different subject groups (Park et al., 2012). Moreover, the supervised/unsupervised feature reduction-based methods offered better classification accuracy than atlas-based methods. An overall improvement of 2–8% over atlas-based methods was seen for different subject groups (86.67%–88.33% for AD/CN (Cho et al., 2012) and 65.22%–71.21% for pMCI/sMCI (Cho et al., 2012)).

3.1.3. Pre-defined regions-based methods—These methods are based on the prior knowledge of the magnitude and spatial pattern of AD that were acquired by studies previously conducted on histological or imaging data (Baron et al., 2001; Frisoni et al., 2002). Generally, features of some of the important regions that have shown to contain discriminatory AD-related information are extracted and used for classification. The datasets and classification accuracies of the studies using these methods are listed in Table 3.

3.1.3.1. Hippocampal features: The hippocampus is amongst the few structures of the medial temporal lobe that undergo severe structural changes in AD (Braak, Braak, 1991). The structural variation between the hippocampus of AD and healthy individuals has been studied intensively (Killiany et al., 2002; Wisse et al., 2014). The geometric properties of the hippocampus have been exploited as useful biomarkers in a few AD and MCI classification studies. For example, the shape of the hippocampus, quantified by spherical harmonics (Gerardin et al., 2009), surface-based anatomic mesh modeling (Li et al., 2007), statistical shape modeling (Shen et al., 2012), and large-deformation diffeomorphic metric mapping and PCA (Wang et al., 2007b), has been shown to be an effective biomarker for AD and MCI classification. The shape and volumetric features of the hippocampus have also been

combined for SVM-based AD conversion prediction (Costafreda et al., 2011). Interestingly, a study demonstrated superiority of hippocampal texture over reduction in its volume for SVM-based prediction of conversion from MCI to AD (Sorensen et al., 2016).

3.1.3.2. Biologically selected features: AD affects brain regions well beyond the hippocampus, such as atrophy of the entorhinal cortex (Dickerson et al., 2001), expansion of the ventricles (Ridha et al., 2008), and volumetric changes in other subcortical nuclei (amygdala, putamen, caudate, and thalamus) (Madsen et al., 1312; Visser et al., 1999). The analysis of structures beyond the hippocampus may not only improve understanding of the spatial pattern of AD, but may also lead to a more precise diagnosis. Therefore, features of these biologically selected regions are sometimes used directly for classification of subjects into normal and diseased classes. For example, Chincarini et al. used the statistical and textural features of the entorhinal cortex, perirhinal cortex, hippocampus, and parahippocampal gyri (Chincarini et al., 2011). The features of each region were analyzed with a random forest classifier to extract the relevant ones, which were subsequently processed with SVM for prediction of AD conversion. Recently, Tang et al. used shape diffeomorphometry of the left and right amygdala, hippocampus, thalamus, caudate, putamen, globus pallidus, and lateral ventricle for prediction of AD conversion using LDA (Tang et al., 2015).

3.2. Functional MRI-based studies

The neurodegenerative process of AD induces changes in functional connectivity between various regions of the brain (Fransson, 2005; Wang et al., 2007a). These alterations are generally measured while the patient is at rest, using resting-state functional MRI (rs-fMRI). Rs-fMRI, in principle, measures the brain activity by quantifying the blood oxygen level-dependent signal, whereby an increased oxygen level is observed in activated regions of the brain due to increased blood flow. Various rs-fMRI studies have reported the existence of resting-state networks, which are characterized by spatially coherent, spontaneous fluctuations in the blood ox gen level-dependent signal and are made up of regional patterns that are commonly involved in brain functions, such as attention, sensory, or default mode processing (Fox, Raichle, 2007; Seeley et al., 2007). A network that is related to AD and has increasingly received attention is the default mode network (DMN) (Greicius et al., 2003; Greicius et al., 2004) also called the ‘task-negative’ network (anti-correlated to ‘task positive’ network) (Fox et al., 2005; Fransson, 2005), since its activity increases in the absence of a task. AD compromises primary brain targets, such as the DMN, by disrupting their functional activity (He et al., 2007; Li et al., 2002), as well as the functional connectivity between primary targets and the remaining parts of the brain (Wang et al., 2007a; Wang et al., 2006). Some studies have reported that functional changes appear well before the changes in structural MRI become evident (Pievani et al., 2011; Teipel et al., 2015). However, those studies are rare, and additional studies should be conducted to test whether functional MRI changes can appear before structural MRI.

The preliminary evidence of disrupted functional connectivity (Li et al., 2002; Wang et al., 2007a; Wang et al., 2006), and its association with AD have led researchers to hypothesize that proper quantification of the functional connectivity across different brain regions can

capture the global distribution of the abnormalities present in AD, and can further aid in AD classification (Chen et al., 2011; Jie et al., 2014). Such quantification involves spatial parcellation of fMRI data according to a structural brain template, and calculation of pair-wise connectivity between the activation in all pairs of regions. The connectivity information, generally defined as the correlation, covariance, or the mutual information between the fMRI time series of the two regions, is stored in an $n \times n$ matrix for each subject, where n is the number of brain regions obtained by parcellation. The connectivity information is then used as input for classification. For example, Chen et al. used Pearson correlation coefficient as connectivity metric for Fisher LDA-based AD and MCI classification (Chen et al., 2011). In addition, Challis et al. used covariance as a connectivity metric for Gaussian-process, logistic regression model-based AD and MCI classification (Challis et al., 2015). They showed that the connectivity strength between the medial structures and temporal and sub-cortical regions best classified MCI, and that the connectivity strength between the frontal areas and the rest of the brain best classified AD. It has also been suggested to develop graphs on connectivity matrcies and compute network measures from the graph instead of using raw connectivity matrices. For example, Jie et al., proposed to extract global topology and local connectivity based features from the graph. The least absolute shrinkage and selection operator was used for feature selection, while multi-kernel SVM was used for MCI classification (Jie et al., 2014). Similarly, Khazaee et al. computed integration and segregation measures from the graph, and used Fisher score for feature selection and SVM for AD classification (Khazaee et al., 2015).

Overall, the functional connectivity-based methods demonstrated good classification results (97.00% for AD/MCI (Challis et al., 2015) and 91.90% for MCI/CN (Jie et al., 2014)).

3.3. Diffusion tensor imaging (DTI)-based studies

AD is associated with loss of brain barriers that restrict water motion, thereby compromising the integrity of WM, and leading to abnormal diffusivity patterns (Xie et al., 2006). DTI is used to analyze water diffusion at the microstructural level of the brain for determining the abnormal diffusion pattern of AD. Voxel-based studies showed that AD and MCI subjects have reduced fractional anisotropy (FA) in multiple posterior WM regions (Medina et al., 2006) and increased mean diffusivity (MD) in the posterior occipital–parietal cortex, and right parietal supramarginal gyrus (Rose et al., 2006). ROI-based studies demonstrated higher MD and/or lower FA in the hippocampus (Fellgiebel et al., 2006; Kantarci et al., 2001; Muller et al., 2005; Muller et al., 2007) and posterior cingulate (Choo et al., 2010; Fellgiebel et al., 2005). Interestingly, a previous study suggested that diffusivity measures of the hippocampus are better predictors of MCI conversion than volume (Fellgiebel et al., 2006). Evidence of abnormal and complex diffusivity patterns has led to the hypothesis that these biomarkers can be used for AD classification using advanced classification framework (Selnes et al., 2013). The studies using DTI-based features, summarized in Table 5, can be further divided into three categories, depending on how features are extracted: i) tractography, ii) connectivity network measures, and iii) discriminative voxel selection.

3.3.1. Tractography-based methods—In this method, the fibers located by means of tractography are clustered into various fiber bundles, based on an anatomical atlas. The

located fibers are reduced to a compact, low-dimensional representation, from which diffusivity measures are calculated for classification. For example, Nir et al. used tractography to locate fibers, then clustered them into 18 fiber bundles based on the 18 regions defined in Johns Hopkins University probabilistic WM tract atlas (Nir et al., 2015). They computed density maps to quantify the number of fibers passing through each voxel, and used the shortest path graph search to reduce fiber bundles to a compact, low-dimensional representation, based on the maximum density path (MDP). These MDPs were registered across subjects, and diffusivity measures of FA and MD, computed along all the MDPs, were used as features for SVM-based AD and MCI classification.

3.3.2. Connectivity network measure-based methods—In this method, DTI images are parcellated into anatomical regions, and several features are calculated from the fibers within these regions. Connectivity networks are developed based on these regions (i.e., features) and a collection of network measures are derived for classification. In this context, Wee et al. parcellated the brain into anatomical regions, and developed connectivity networks based on the regional features of fiber count, averages of on-fiber FA, MD, and three principal diffusivities (Wee et al., 2011). Clustering coefficients of all the regions, computed for all the networks, were used as features. The feature set was reduced by determining the Pearson correlation coefficient, and SVM-RFE was used for MCI classification. Recently, Prasad et al. adopted the same methodology, and developed two connectivity networks based on regional features of fiber count, and flow along the fibers (Prasad et al., 2015). Raw connectivity matrices and various other network measures, such as global efficiency, transitivity, path-length, modularity, radius, and diameter were used for SVM-based classification of early and late MCI subjects.

3.3.3. Discriminative voxel selection-based methods—In this method, discriminative voxels are selected to reduce the dimensionality of DTI data, and diffusion measures of selected voxels are used as features for classification. Dyrba et al. adopted this approach in two of their studies. In the first study, they used PCA and an entropy-based information gain criterion for selecting discriminative voxels, and used diffusion measures of the FA, MD and mode of anisotropy of the selected voxels as features for SVM-based AD classification (Dyrba et al., 2013). In the subsequent study (Dyrba et al., 2015a), they used the same classification framework for classifying MCI subjects, stratified by their positive or negative amyloid burden.

3.4. Positron-emission tomography (PET)-based studies

The characteristic patterns of glucose metabolism on brain FDG-PET and of amyloid deposition on amyloid PET can help in differentiating AD from healthy individuals. An association between AD and hypometabolism was found in several brain regions, such as the parieto–`temporal and posterior cingulate cortices (Mosconi et al., 2008), and hippocampus (Mosconi et al., 2005). Similarly, AD subjects compared to healthy individuals have shown higher amyloid burden in overall cortex and all cortical regions (precuneus, anterior and posterior cingulate, and frontal median, temporal, parietal and occipital cortex) (Camus et al., 2012). These evidences of the association of hypometabolism and amyloid burden with

AD encouraged the use of FDG-PET and amyloid PET as a suitable biomarker for AD classification.

3.4.1. FDG-PET—Recently, there has been a growing interest in using the cerebral glucose metabolism rate for AD classification and prediction of conversion from MCI to AD (Gray et al., 2012; Toussaint et al., 2012). Four main methods that utilize the cerebral glucose metabolism rate are considered here: voxels as feature (VAF)-based, discriminative voxel selection-based, atlas-based, and projection-based methods. Table 6 lists the studies using these methods, and their datasets and corresponding classification performance.

3.4.1.1. VAF-based methods: This set of methods, similar to those discussed for structural MRI in section 3.1, utilizes the intensity value of all the voxels of an input PET scan. Hinrichs et al. adopted this approach for AD classification using LPBM (Hinrichs et al., 2009).

3.4.1.2. Discriminative voxel selection-based methods: The prominent goal of these methods is to simultaneously select the informative voxels (features) used in VAF-based methods. For example, Cabral et al. used this method to investigate the prediction of AD conversion based on FDG-PET images at various time-points (Cabral et al., 2015). Discriminative voxels of the images were selected using mutual information criterion, and SVM and Gaussian naive bayes were used for classification. The classification based on stage-specific categories of pMCIs demonstrated better predictive accuracy than did the overall classification of pMCI and sMCI, a result also demonstrated earlier by (Eskildsen et al., 2013).

3.4.1.3. Atlas-based methods: Several FDG-PET based AD classification studies are based on parcellation of PET images into different anatomical regions, utilizing pre-defined structural atlases. Pagani et al. used the average regional intensity and inter-hemispheric symmetry between the parcellated regions as features for SVM-based classification (Pagani et al., 2015). Similarly, Gray et al. used the average regional intensities of baseline and 12-months follow-up scans, and the difference of intensity between these two time-points as features for SVM-based classification (Gray et al., 2012). Accuracy increased by a factor of 1–2% when using both the longitudinal and cross-sectional features in this study.

3.4.1.4. Projection-based methods: Projection-based methods reduce the dimensionality of features by projecting the higher-dimensional feature space into a lower-dimensional space, where the significance of each feature, with respect to the problem at hand, can be measured in terms of its variance. Thus, a subset of features with relatively larger variance may be selected for further inspection. In this context, Padilla et al. applied non-negative matrix factorization projections to input images (Padilla et al., 2012). They then selected several subsets of these projections, and classified those using SVM. Ultimately, the classification results obtained from several projections were combined to obtain a final prediction. The authors showed that this method achieved a 17% improvement in classification accuracy as compared to a VAF-based method for the same dataset.

3.4.2. Amyloid-PET—*In vivo* measurements of the cerebral am loid burden (β-amyloid) may be clinically useful in the management of patients with cognitive impairment who are being evaluated for possible AD. Several radioligands are used for measuring amyloid burden, such as the ^{11}C-Pittsburgh Compound B (^{11}C-PIB), and ^{18}F-labelled amyloid-PET tracers, including florbetapir, flutemetamol, and florbetaben. Some group analysis-based studies have revealed differences in amyloid burden in various brain regions in different subject groups; higher amyloid burden has been found in MCI (Klunk et al., 2004) and AD subjects (Camus et al., 2012) than in healthy individuals, and in subjects with pMCI than in those with sMCI (Koivunen et al., 2011; Okello et al., 2009). However, very little attention has been paid to quantification of *plaque* levels in different brain regions and its use in AD classification. Vandenberghe et al. proposed one such classification framework in which they used the intensity values of all the voxels of ^{18}F-flutemetamol PET scans as features for SVM-based classification of AD versus healthy individuals, and pMCI versus sMCI subjects (Vandenberghe et al., 2013).

3.5. Multimodal studies

Several biomarkers have shown association with AD, including proteins measured in the CSF (Melah et al., 2016), brain atrophy, particularly in the hippocampus (Frisoni et al., 2002) and posterior cingulate gyrus (Baron et al., 2001), measured through structural MRI, and hypometabolism, associated with AD in the temporal and parietal lobe, as well as in the posterior cingulate cortex, measured via functional imaging (Herholz et al., 2002; Langbaum et al., 2009). In addition, AD brains demonstrate the formation of insoluble β amyloid *plaques* and neurofibrillary tangles (Jagust, 2013), and it has been suggested that the quantity of β am loid can be related to the disease stage (Murpy, LeVine III, 2010).

These biomarkers yield complementary information, i.e., different modalities capture disease information from different perspectives, thereby improving understanding of the disease pattern over that presented by one modality. Classification frameworks facilitate exploitation of the complementary information obtained from multiple modalities. A top-level breakdown of two classification frameworks, straightforward feature concatenation and specialized fusion frameworks, used to exploit multimodal data for AD classification is shown in Figure 3. The terms $F_1, F_2, \ldots, F_{n-1}, F_n$ are the feature sets extracted from $1, 2, \ldots, n-1, n$ modalities or other biomarkers, respectively. The symbols $C_1, C_2, \ldots, C_{n-1}, C_n$ represent the classification algorithms trained on feature sets $1, 2, \ldots n-1, n$, respectively. Table 7 summarizes the results and the corresponding datasets for the studies using these frameworks.

3.5.1. Straightforward feature concatenation—The simplest method for exploiting the complementary information provided by multiple modalities is concatenation of the features of these modalities into a single feature vector and training a classifier on that vector.

Structural MRI is a key component of these studies, and its features are combined with features extracted from various other modalities to improve classification. In this context, various authors have combined structural MRI-based features with the features calculated

from DTI and functional MRI. In earlier studies, the regional volumetric measures, calculated from structural MRI, and FA, calculated from WM tracts, have been combined for SVM-based MCI and AD classification (Cui et al., 2012; Li et al., 2014a). Among recent studies, Tang et al. used volumetric, shape, and diffusion features of the hippocampus and amygdala for AD classification (Tang et al., 2016). They used PCA and tudent's t-test for reducing the feature set, and LDA and SVM for classification. Similarly, Schouten et al. combined regional volumetric measures, diffusion measures, and correlation measures amongst all brain regions calculated from functional MRI (Schouten et al., 2016). They employed logistic elastic net (Zou, Hastie, 2005) for classification.

The combination of structural MRI with demographics, cognitive tests, and genetic data has also been explored in a few studies. For instance, Vemuri et al. combined GM, WM, and CSF density maps with age, gender, and APOE genotype for SVM-based AD classification (Vemuri et al., 2008). Zhang et al. used GM density maps, intracranial volume, atlas-scaling factor, normalized whole brain volume, and age as features for AD and MCI classification (Zhang et al., 2014). The feature set was classified by using a kernel SVM decision-tree (a variant of the SVM decision-tree). Recently, Moradi et al. used GM density maps, age, and cognitive tests as features, and employed classification algorithms such as low-density separation and random forest for AD conversion prediction (Moradi et al., 2015).

The combination of structural MRI features with PET and CSF biomarkers is another dimension. In earlier studies, Fan et al. combined regional volumetric measures, and regional cerebral blood flow for MCI classification using SVM (Fan et al., 2008b). In addition to this, the SPARE-AD index (Davatzikos et al., 2009) was combined with CSF biomarkers (Davatzikos et al., 2011) to predict conversion from MCI to AD using SVM. Further, Dukart et al. first used FDG-PET and GM density values of all the voxels of selected ROIs (Dukart et al., 2011a) and later the average of all the voxels of selected ROIs for SVM-based AD classification (Dukart et al., 2013). In recent studies, Zhu et al. combined regional GM volume, regional average PET intensity, and CSF biomarkers as features, and proposed a matrix-similarity-based loss function for better classification using SVM (Zhu et al., 2014). Similarly, Zheng et al. used regional thickness measures, regional correlative measures (calculated from thickness measures) and APOE genotype for SVM-based AD classification, and AD prediction conversion (Zheng et al., 2015). Apostolova et al. used hippocampal volume and CSF biomarkers for SVM-based AD and MCI classification (Apostolova et al., 2014).

The combination of structural MRI, PET, and CSF biomarkers together with genetic data and neuropsychological status exam scores has also been common. For example, SPARE-AD was combined with cognitive scores, APOE genotype, and CSF biomarkers (Da et al., 2014) to predict conversion from MCI to AD. Similarly, Kohannim et al. combined hippocampal, ventricular, and temporal lobe volumes, FDG-PET numeric summary, CSF biomarkers, APOE genotype, age, sex, and body mass index for SVM-based AD and MCI classification (Kohannim et al., 2010). In addition, Cui et al. combined average regional cortical thickness, standard deviation of thickness, average regional surface area and cortical volume from structural MRI with CSF biomarkers and neuropsychological status exam

scores (Cui et al., 2011). Features were reduced using mRMR, and SVM was used for classification.

3.5.2. Specialized fusion strategies—While the simplicity of straightforward concatenation may be considered desirable, the method suffers from a major pitfall: because it treats all features as equivalent, it provides no way to account for the different natures of features extracted from different modalities (Hinrichs et al., 2011; Liu et al., 2014). For instance, where one modality has many more features than another (or has variation on a much larger scale), classification algorithms trained on concatenated features may produce prediction models that effectively ignore the other modalities. Specialized fusion strategies can be used to ensure that the complementary information found across all modalities is still used. These strategies may either combine the results of classification rules trained on the individual modalities (Dai et al., 2012) or use special combination rules to combine features before training (Dyrba et al., 2015b; Zhang, Shen, 2012; Zhang et al., 2011).

These strategies were employed to utilize the complementary information of features extracted from structural MRI, rs-fMRI, and DTI for AD classification. Dai et al. used regional GM volumetric measures, and functional measures (amplitude of low-frequenc fluctuations, regional homogeneit, and regional functional connectivity strength) as features (Dai et al., 2012). They trained separate maximum uncertainty LDA classifiers on the structural and functional measures, and combined the output of the classifiers via weighted voting. Recently, Dyrba et al. used regional GM volumetric measures, average tract intensity for FA, MD, and mode of anisotropy, and network measures of weighted local clustering coefficient and the shortest weighted path-length calculated from rs-fMRI as features (Dyrba et al., 2015b). They adopted multi-kernel SVM for AD classification.

In addition, the features extracted from structural MRI and PET images, as well as CSF biomarkers were also fused using these strategies. In this context, regional GM volume, regional average FDG-PET intensity, and CSF biomarkers were used as features for AD and MCI classification along with multi-kernel learning (Zhang, Shen, 2012; Zhang et al., 2011) and multi-task learning (Yu et al., 2016). The same features were also used for prediction of conversion from MCI to AD by using domain transfer learning (Cheng et al., 2015b) and semi-supervised multimodal manifold-regularized transfer learning (Cheng et al., 2015a). Recently, Liu et al. used CSF biomarkers, and shape measures of hippocampus and GM volume in atlas-defined ROIs for AD classification using multi-kernel SVM (Liu et al., 2014). Xu et al. proposed using the GM volume, regional average intensity from FDG-PET and Florbetapir images as features (Xu et al., 2015). They assigned different weights to the features of different modalities for classification of AD versus MCI subjects, and pMCI versus sMCI subjects using a sparse representation-based classification method. Zu et al. used regional GM volume, and the regional average FDG-PET intensity for AD and MCI classification, and used multi-kernel SVM for classification (Zu et al., 2015).

Furthermore, these strategies were also used to exploit the features of structural MRI, PET, and CSF biomarkers together with genetic data and neuropsychological status exam scores. In earlier studies, Hinrichs et al. used structural MRI-based density maps, FDG-PET intensities, CSF biomarkers, APOE genotype, and neuropsychological status exam scores as

features for multi-kernel SVM-based AD classification, and AD conversion prediction (Hinrichs et al., 2011). Young et al. used regional GM volume, regional average FDG-PET intensity, and CSF biomarkers along with APOE genotype, and employed Gaussian process classifier to train multiple kernels on AD and CN subjects, and test on pMCI and sMCI subjects (Young et al., 2013). Also, Gray et al. trained random forest classifiers on regional GM volumetric measures, CSF biomarkers, voxel-based FDG-PET intensities, and APOE genotype for AD and MCI classification (Gray et al., 2013). Casanova et al. combined SPARE-AD index with GM, WM and CSF density maps, total hippocampal volume, regional volumetric and cortical thickness measures, and cognitive scores for AD classification and prediction of conversion from MCI to AD using large-scale RLR (Casanova et al., 2013). In recent studies, Korolev et al. used cortical and subcortical volumes, mean cortical thickness, surface area, and curvature from structural MRI, clinical measures and plasma measures for AD conversion prediction. They used mutual information criterion for feature selection and probabilistic multi-kernel learning for classification (Korolev et al., 2016). Similarly, Clark et al. used cognitive scores, cortical thickness measures, hippocampal and ventricular volume along with age, sex, and education. They selected features using random forest, and used an ensemble of random forests of conditional trees, SVM, naive Bayes, and multilayer perceptrons for classification. (Clark et al., 2016).

Overall, the multimodal techniques under this category have demonstrated varied improvement over single modalities, ranging from 1 to 7%. Almost all the methods demonstrated some improvement (Dyrba et al., 2015b), however, no improvement was observed using multimodality data as compared to DTI measures alone.

4. Discussion

Recent advances in neuroimaging research suggest that AD pathology can be detected preclinically (Perrin et al., 2009). Consequently, an important body of research has been devoted to the neuroimaging-based AD/MCI classification and AD conversion prediction using various neuroimaging modalities, such as structural MRI, functional MRI, DTI, and PET. In this review, we presented only those studies that used appropriate cross-validation strategies to assess the performance of their classification frameworks. Among the various neuroimaging modalities, structural MRI was the most frequently used, likely due to its widespread availability. The second most common method was the combination of features from one or more modalities with data levels such as genetic data, cognitive scores, and CSF biomarkers. Research based on features extracted from FDG-PET, amyloid-PET, DTI, and functional MRI were less common. The main objective in most of the studies reviewed here was the production and selection of AD-related inherent features from high-dimensional raw neuroimaging data. Therefore, we grouped the classification studies appertaining to each modality according to feature extraction methods. Brain atrophy was most often quantified via tissue density maps (Casanova et al., 2011; Kloppel et al., 2008), cortical/subcortical thickness measurements (Desikan et al., 2009; Wee et al., 2013), and geometric measures of hippocampus (Costafreda et al., 2011; Gerardin et al., 2009) from structural MRI. Connectivity networks developed on top of the functional strength (Chen et al., 2011; Koch et al., 2012) and diffusion measures (Prasad et al., 2015; Wee et al., 2011) of parcellated

brain regions were the most common feature extraction methods in functional MRI and DTI, respectively. Similarly, the cerebral glucose metabolic rate measured in parcellated brain regions was common in FDG-PET (Gray et al., 2012; Pagani et al., 2015). These automated feature extraction methods generate a high-dimensional data for further analysis. A wide variety of sophisticated and well-established supervised classification algorithms such as SVM and LDA have been applied on extracted neuroimaging features for AD/MCI classification or AD prediction conversion in different studies.

The main advantage of applying classification algorithms on neuroimaging data is the potential use for detecting AD at the prodromal stages, well even before clinical manifestation (Misra et al., 2009; Park et al., 2013), demonstrating the probable use in routine clinical settings in the future. Among the wide range of classification algorithms, SVM was the most frequent for AD classification (Apostolova et al., 2014; Cui et al., 2012; Kohannim et al., 2010; Padilla et al., 2012). Multi-kernel learning, which are an extension of ordinary kernel-based classification algorithms, were also increasingly used in AD classification (Dyrba et al., 2015b; Liu et al., 2014; Zu et al., 2015). Other less common classification algorithms used in AD research were LDA (Lillemark et al., 2014; Tang et al., 2015), orthogonal partial least square regression (Westman et al., 2011), random forest (Moradi et al., 2015), regularization-based methods (Casanova et al., 2011), voting-based ensemble methods (Liu et al., 2015), kernel SVM decision-tree (Zhang et al., 2014), and LPBM (Hinrichs et al., 2009). While SVM could have the advantage of achieving high classification accuracy with small training sample size compared to other classification algorithms such as neural networks (Shao, Lunetta, 2012), they might have the disadvantage of the need for parameter tuning (Chapelle et al., 2002). For neuroimaging AD classification it remains important to conduct studies comparing between diverse classification algorithms thoroughly, as only limited number of studies have been conducted so far (Khondoker et al., 2016; Lehmann et al., 2007).

The feature extraction methods summarized here are influenced by several factors that vary across studies. One factor is spatial smoothing of structural MRI and FDG-PET, which is generally performed to account for noise (i.e. registration errors). Usually, Gaussian smoothing of full-width half-max is used for denoising. It is important to note that too small kernel size might lead to missing the many regions that might present group differences. Conversely, too large kernel may blur image features in regions that display group differences from the rest of the regions. An optimal solution has yet to be achieved as the kernel size is chosen either ad hoc or empirically. A majority of the reviewed studies used a Gaussian kernel of 8 mm for both structural MRI (Misra et al., 2009; Moradi et al., 2015) and FDG-PET (Gray et al., 2013; Pagani et al., 2015; Zhang et al., 2011; Zhu et al., 2014). However, kernels of other sizes, such as 10 mm for structural MRI (Dai et al., 2012; Plant et al., 2010) and 15 mm for FDG-PET (Fan et al., 2008b), were rarely used. An additional factor that influences the atlas-based methods is the selection of the atlas itself. Atlas-based parcellation using a pre-defined anatomical brain atlas is a methodologically simple and computationally tractable feature extraction method, with general versatility (Ota et al., 2015; Zhang et al., 2011). However, the choice of atlas will have an effect on classification performance. It has been shown that features extracted based on different anatomical parcellations lead to differences in classification performance under similar experimental

conditions (Ota et al., 2014, 2015). These differences in classification performances may be associated with changes in parcellation between atlases, for example, the cerebellum region. The LONI probabilistic brain atlas considers the cerebellum as one single region, whereas the AAL atlas finely parcellates the cerebellum into 26 smaller regions.

In addition to feature extraction and classification, feature selection is also important for identifying distinguishing features. Selection of appropriate features not only removes the non-informative signal, but also reduces the computational time involved in classification. Two widely adopted methods for feature selection are biologically informed and automated feature selection. The former relies on prior biological knowledge about the discriminating ability of certain regions, generally obtained from existing literature, whereas the latter selects features based on general data characteristics, without prior knowledge. Among the automated methods, various ranking-based methods, such as t-tests (Tang et al., 2016; Wee et al., 2013) and earson's correlation coefficient test (Davatzikos et al., 2008; Wee et al., 2011), wrapper-based methods, a combination of ranking and wrapper-based methods, such as mRMR (Wee et al., 2013), and embedded methods, such as elastic net regression, were used in the reviewed studies, and improved the classification performance. It is feasible that variations in feature-selection methods will lead to differences in AD classification performance. It has been suggested that automated feature selection will not improve classification accuracy as compared to biologically informed feature selection, driven by prior biological knowledge of regions typically affected by AD, such as the hippocampus, amygdala, thalamus, and caudate (Chu et al., 2012). Similar results were observed in the Pittsburgh Brain Activity Interpretation Competition, wherein the team applying prior biological knowledge for feature selection (Chu et al., 2011) outperformed the teams using automated feature selection. In addition, the winning method (Sørensen et al., 2014) in the recent CADDementia 2015 challenge (Bron et al., 2015) was also based on biologically informed feature selection.

4.1. AD classification studies comparison

The key components of each classification study, such as prediction accuracy, study population, and feature types were summarized in table format in this review. It should be emphasized that these tables are meant to provide a glance to each individual study and not for comparative purposes. Frequently throughout reviewing these studies, authors stated that their proposed classification framework was superior to existing ones solely on the basis of the achieved accuracy. However, we believe that considering the number of factors involved in each study, summarized below, it is difficult to compare these studies directly and therefore to draw general conclusions about the state of the field as a whole.

Length of follow-up period—The length of the follow-up period for defining MCI conversion also varied from a minimum of 6 months to a maximum of 36 months across different studies. It is well-known that the level of neurodegeneration, and hence, the rate of AD prediction increases as the MCI subjects progress on a continuum from the CN state to the AD state (Cabral et al., 2015; Eskildsen et al., 2013). Therefore, we believe that the prediction performance of various studies cannot be compared directly, considering the different lengths of the follow-up periods.

Study population—Baseline characteristics of the study population, such as gender, age, genotype (APOE), and education are considered to be confounding factors in AD classification. These factors may have a profound effect on key features extracted from the neuroimaging modalities and therefore on the resultant classification accuracies. AD classification studies in general differ in how they deal with the confounding factors, i.e., the number of confounding factors that need to be considered, and how they are considered, etc. In the past, confounding factors have mainly been dealt with by matching the subjects in different groups according to the factors or by using confounding factors as covariates in a statistical model, in order to remove their effect from the model. However, increasing attention has been given to this issue in recent years, and several automated methods have now been proposed to control for the effects of confounding factors (Dukart et al., 2011b; Li et al., 2011).

Degree of impairment—An objective, final diagnosis of AD can only be made through autopsy and therefore is rarely used (Kloppel et al., 2008). Even then, the disease stage at autopsy can be very different from the disease stage determined by scanning. Alternatively, clinical diagnostic criteria for AD (McKhann et al., 2011) and MCI (Petersen, 2004) are used in practice as a reference standard for evaluation. The MCI diagnosis based on these criteria leads to a clinically heterogeneous mix of more and less impaired patients, where each patient presents a disease stage on a continuum from CN to AD (Misra et al., 2009). A more severely impaired MCI group, when MCI is used as one diagnostic entity, may show larger structural differences from healthy individuals, leading to potentially higher classification accuracies.

Evaluation metrics—Classification performance in some of the studies reviewed here was only reported in terms of classification accuracy. The measure of classification accuracy by itself could be uninformative in unbalanced datasets and cannot be used for comparison. For instance, a 90% classification accuracy in a dataset of 90 diseased and 10 healthy individuals does not convey any information, since a biased classification algorithm that classifies all the subjects as diseased can also lead to a 90% classification accuracy. Therefore, we believe that balanced accuracy; sensitivity/specificity or precision/recall, along with the area under a receiver-operating-characteristic curve should be reported for direct comparison of results.

Factors affecting the performance of classification algorithms—The expected performance of a classification algorithm is defined by two factors. The first is the number of subjects in the training set and the second is the relative proportion of subjects from each class present in the test set. Mostly the first factor is determined by the sample size available for training, and by the cross-validation strategy used in the experiment. The larger the number of subjects in the training set, the better the generalizability of the classification algorithm. The second factor could affect the classification accuracy as depending on class relative proportion, the sensitivity and specificity of a classification algorithm could differ. This factor can be easier to fix, as stratification is becoming increasingly common, and subjects of different classes are selected based on matched demographic characteristics.

The choice of the split-in-train-test or k-fold cross-validation strategies, adopted in the classification framework, influences the statistical significance of the classification accuracy (Mendelson et al., 2014), which can be calculated via binomial or permutation tests (Noirhomme et al., 2014). The split-in-train-test assumes independence between the training and test sets, which is the key to binomial tests. However, the split-in-train-test can generally be limited in some medical applications where classification algorithms are trained on small number of subjects. Consequently, k-fold cross-validation is more commonly applied. K-fold cross-validation does not hold the independence assumption, as the training subjects in different iterations could overlap; therefore, permutation tests are feasible for evaluating the statistical significance of k-fold cross-validation strategies (Noirhomme et al., 2014). It has been suggested that 10-fold or 5-fold cross-validation should be used to establish a trade-off between bias and variance (Lemm et al., 2011). Furthermore, it has been suggested that permutation tests should be used along with cross-validation, especially when dealing with a small sample size (Noirhomme et al., 2014).

4.2. Challenges with AD classification studies

Generalization ability—A critical challenge underlying the clinical use of AD classification frameworks is the ability of predictive models that allow good generalization to new patient data. Ideally, the models should be able to perform well regardless of the variability of imaging protocols, scanners (Abdulkadir et al., 2011) and demographics, and should be free of double-dipping, a phenomenon very common in older studies. The term double-dipping, or circular analysis, refers to the use of test subjects in any part of the training process, such as selection of features and training of classification algorithm, and may lead to over-fitted classification (Kriegeskorte et al., 2009). To conduct a fair validation, one should avoid double-dipping by excluding the test subjects used in the subsequent validation of the classification algorithms from the process of feature selection and training of the classification algorithm. Double-dipping was quite common in feature selection in older studies (Querbes et al., 2009; Wolz et al., 2011); however, it has become less common as its effects became clearer. In order to encourage the development of classification frameworks that are generalizable to new datasets, the neuroimaging community has organized more AD classification challenges, in which different researchers attempt to solve the classification problem by leveraging the current state-of-the-art techniques on publicly available datasets.

For example, the purpose of CADDementia 2015 challenge (Bron et al., 2015) was to measure the generalizability of structural MRI-based classification studies on unseen subjects, where the best performing study yielded an area under the receiver-operating-characteristic curve of 78.8%. Similarly, the main aim of the DREAM 2016 challenge (Allen et al., 2016) was to identify accurate biomarkers of cognitive decline for advancing early diagnosis. These challenges not only help to determine the generalizability of any study, but also enable fair performance comparison of different studies on the same dataset, which would otherwise not be possible due to different experimental conditions across distinct studies. Furthermore, efforts have been made to standardize comparison of various studies on the same dataset, such as in the study of Cuingnet et al., in which the authors evaluated the performances of 10 studies using the ADNI dataset (Cuingnet et al., 2011).

Sample size—It is generally believed that smaller datasets do not capture the full spectrum of heterogeneity among different classes and therefore, may be less generalizable on unseen patient data, whereas opposite is true for larger datasets. Nonetheless, quite different results were seen in CADDementia 2015 challenge (Bron et al., 2015), where the studies training classification algorithms on larger training sets (Abdulkadir et al., 2014; Eskildsen et al., 2014) did not perform better than the studies training classification algorithms on relatively smaller datasets. Therefore, the minimum sample size required for training a generalizable classification model remains debatable.

Reproducibility—Disregard of the appropriate packaging of classification frameworks, where numerical solutions could lead to different experimentation conditions, and the use of local datasets or a subset of a larger public dataset without providing detailed subject level identification, are the main factors hindering the reproducibility of existing results and comparison between study findings. We highly encourage reporting the results derived from public datasets, and appropriate listing of the subjects, as had been done in a few studies (Moradi et al., 2015; Zhang et al., 2011). We also recommend that authors attend to the proper packaging and availability of their code, particularly in cases where sophisticated feature extraction and classification algorithms have been used, as this can markedly improve reproducibility.

AD heterogeneity—The heterogeneity of AD necessitates a definition of distinct clinicopathological subtypes of AD. While AD has been stereotypically defined using the Braak stages, atypical AD cases do not fit into this scheme. For example, a recent study has shown that hippocampal sparing and limbic-predominant AD subtypes might account for about 25% of AD cases (Murray et al., 2011). Simplistic measures, such as the ratio of hippocampal to cortical volumes, showed a high discrimination ability between the subtypes (Whitwell et al., 2012). Other studies used clustering-based approaches for defining the AD pattern; a recent study made use of cortical thickness clustering and showed that AD in the earlier stages can be categorized into various anatomical subtypes, with distinct clinical features (Noh et al., 2014). By including additional biomarkers, such as cerebrospinal fluid and serum biomarkers, four clusters emerged with distinct biomarker patterns, the first of which was biologically similar to healthy individuals and which rarely converted to AD (Nettiksimmons et al., 2014). We believe that the heterogeneity of AD patterns has been widely ignored in the existing AD classification studies, and more attention should be paid to this line of research in future. The development of tools that can deal with heterogeneous imaging patterns is important and should become an area of focus (Dong et al., 2016a; Dong et al., 2016b; Varol et al., 2017). It is likely that systematic quantification of heterogeneity is critical for developing effective personalized diagnostic and predictive tools using machine learning.

4.3. Conclusion and future directions

Neuroimaging-based classification of AD and MCI has increasingly been reported in the literature over the past decade, as a means to derive individual biomarkers of these conditions. The ultimate goal of AD classification is to generate an individual diagnosis using a single MRI scan by applying classification models already trained on a large pool of

diseased and healthy individuals, and to predict future progression at earlier disease stages. Several neuroimaging modalities, as discussed in this review, including structural and functional MRI, DTI, FDG-PET, and amyloid-PET, have shown characteristic alterations in the brains of AD and MCI patients that can help rule-in the pathophysiological process of AD. No single neuroimaging modality can be sufficient, as each has complementary merits and limitations. Combining information from multiple modalities has improved the classification performance of AD/MCI and AD conversion prediction. In addition, the combination of features extracted from neuroimaging modalities with demographics, cognitive test scores, CSF biomarkers, and genetic data were also effective in achieving accurate classification. However, there is a great need for validation of these markers in clinical settings, along with their validation in databases comprising highly preselected subjects, which significantly differ from that seen in the clinic. Some challenges faced by the researchers in the field of AD classification, such as high dimensions of raw neuroimaging data, smaller sample sizes, generalizability, and heterogeneity in AD, make it difficult to derive a more precise classification. However, the use of neuroimaging for AD classification remains highly promising, as many of the aforementioned challenges can be addressed.

The potential consideration of classification frameworks in clinical practice has largely driven the development of machine-learning tools that can integrate several imaging features and make predictions on an individual basis. This line of research is likely to become a focus-point in the upcoming decade. In addition, multimodal approaches that seek to find patterns of neurodegeneration across different types of images that form distinctive imaging signatures of the stages of AD, and consensus-based approaches, which tend to improve classification by combining the output of multiple classification algorithms, are also gaining increasing attention. Biologically informed feature selection, and characterization of heterogeneity of AD are also important lines of research that are likely to be emphasized in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by NIH grant R01AG14971. We would also like to thank the anonymous reviewers, whose detailed and thoughtful comments helped us significantly improve the paper.

Abbreviations (Acronym Abbreviation)

AAL	Automated anatomical labeling
AD	Alzheimer’s disease
ADNI	Alzheimer’s disease neuroimage initiative
AUC	Area under a receiver-operating-characteristic curve
CN	Cognitively normal

CSF	Cerebrospinal fluid
DMAF	Density maps as features
DMN	Default mode network
DTI	Diffusion tensor imaging
FDG-PET	Fluorodeoxyglucose positron emission tomography
fMRI	Functional magnetic resonance imaging
GM	Gray matter
LDA	Linear discriminant analysis
LPBM	Linear program boosting method
LR	Logistic regression
MCI	Mild cognitive impairment
MRI	Magnetic resonance imaging
mRMR	Minimum-redundancy and maximum-relevance
OASIS	Open access series of imaging initiatives
pMCI	Progressive mild cognitive impairment
RAVENS	Regional analysis of volumes examined in normalized space
ROI	Region of interest
RLR	Regularized logistic regression
RVM	Relevance vector machines
rs-fMRI	Resting state functional magnetic resonance imaging
sMCI	Stable mild cognitive impairment
sMRI	Structural magnetic resonance imaging
SPARE-AD	Spatial pattern of abnormality for recognition of early Alzheimer’s disease
SVM	Support vector machines
SVM-RFE	Support vector machines-recursive feature elimination
VAF	Voxels as features
VBM	Voxel based morphometry
WM	White matter

References

Abdulkadir A, Mortamet B, Vemuri P, Jack CR Jr, Krueger G, Kloppel S. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. Neuroimage. 2011; 58:785–792. [PubMed: 21708272]

Abdulkadir A, Peter J, Brox T, Ronneberger O, Klöppel S. Voxel-based multi-class classification of AD, MCI, and elderly controls: Blind evaluation on an independent test set. Proc MICCAI workshop challenge on computer-aided diagnosis of dementia based on structural MRI data. 2014:8–15.

Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B. How early can we predict Alzheimer's disease using computational anatomy? Neurobiology of aging. 2013; 34:2815–2826. [PubMed: 23890839]

Agosta F, Pievani M, Geroldi C, Copetti M, Frisoni GB, Filippi M. Resting state fMRI in Alzheimer's disease: beyond the default mode network. Neurobiol Aging. 2012; 33:1564–1578. [PubMed: 21813210]

Allen GI, Amoroso N, Anghel C, Balagurusamy V, Bare CJ, Beaton D, Bellotti R, Bennett DA, Boehme KL, Boutros PC, et al. Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. Alzheimers Dement. 2016; 12:645–653. [PubMed: 27079753]

Apostolova LG, Hwang KS, Kohannim O, Avila D, Elashoff D, Jack CR Jr, Shaw L, Trojanowski JQ, Weiner MW, Thompson PM. ApoE4 effects on automated diagnostic classifiers for mild cognitive impairment and Alzheimer's disease. Neuroimage Clin. 2014; 4:461–472. [PubMed: 24634832]

Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage. 2017; 145:137–165. [PubMed: 27012503]

Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statistics surveys. 2010; 4:40–79.

Ashburner J, Friston KJ. Voxel-based morphometry—the methods. Neuroimage. 2000; 11:805–821. [PubMed: 10860804]

Bakkour A, Morris JC, Dickerson BC. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. Neurology. 2009; 72:1048–1055. [PubMed: 19109536]

Baron JC, Chetelat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. Neuroimage. 2001; 14:298–309. [PubMed: 11467904]

Beheshti I, Demirel H. Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease. Comput Biol Med. 2015; 64:208–216. [PubMed: 26226415]

Binnewijzend MA, Schoonheim MM, Sanz-Arigita E, Wink AM, van der Flier WM, Tolboom N, Adriaanse SM, Damoiseaux JS, Scheltens P, van Berckel BN, et al. Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging. 2012; 33:2018–2028. [PubMed: 21862179]

Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 1991; 82:239–259. [PubMed: 1759558]

Bron EE, Smits M, van der Flier WM, Vrenken H, Barkhof F, Scheltens P, Papma JM, Steketee RM, Mendez Orellana C, Meijboom R, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. Neuroimage. 2015; 111:562–579. [PubMed: 25652394]

Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 2007; 3:186–191. [PubMed: 19595937]

Buckner RL. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. Neuron. 2004; 44:195–208. [PubMed: 15450170]

Busatto GF, Garrido GE, Almeida OP, Castro CC, Camargo CH, Cid CG, Buchpiguel CA, Furuie S, Bottino CM. A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. Neurobiol Aging. 2003; 24:221–231. [PubMed: 12498956]

Cabral C, Morgado PM, Campos Costa D, Silveira M. Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. Comput Biol Med. 2015; 58:101–109. [PubMed: 25625698]

Camus V, Payoux P, Barre L, Desgranges B, Voisin T, Tauber C, La Joie R, Tafani M, Hommet C, Chetelat G, et al. Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment. Eur J Nucl Med Mol Imaging. 2012; 39:621–631. [PubMed: 22252372]

Casanova R, Hsu FC, Sink KM, Rapp SR, Williamson JD, Resnick SM, Espeland MA. Alzheimer's disease risk assessment using large-scale machine learning methods. PLoS One. 2013; 8:e77949. [PubMed: 24250789]

Casanova R, Whitlow CT, Wagner B, Williamson J, Shumaker SA, Maldjian JA, Espeland MA. High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. Front Neuroinform. 2011; 5:22. [PubMed: 22016732]

Challis E, Hurley P, Serra L, Bozzali M, Oliver S, Cercignani M. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. Neuroimage. 2015; 112:232–243. [PubMed: 25731993]

Chan D, Janssen JC, Whitwell JL, Watt HC, Jenkins R, Frost C, Rossor MN, Fox NC. Change in rates of cerebral atrophy over time in early-onset Alzheimer's disease: longitudinal MRI study. Lancet. 2003; 362:1121–1122. [PubMed: 14550701]

Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing Multiple Parameters for Support Vector Machines. Machine learning. 2002; 46:131–159.

Chen G, Ward BD, Xie C, Li W, Wu Z, Jones JL, Franczak M, Antuono P, Li SJ. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. Radiology. 2011; 259:213–221. [PubMed: 21248238]

Cheng B, Liu M, Suk HI, Shen D, Zhang D. Multimodal manifold-regularized transfer learning for MCI conversion prediction. Brain Imaging Behav. 2015a; 9:913–926. [PubMed: 25702248]

Cheng B, Liu M, Zhang D, Munsell BC, Shen D. Domain Transfer Learning for MCI Conversion Prediction. IEEE Trans Biomed Eng. 2015b; 62:1805–1817. [PubMed: 25751861]

Chetelat G, Baron JC. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. Neuroimage. 2003; 18:525–541. [PubMed: 12595205]

Chiang HS, Pao SC. An EEG-Based Fuzzy Probability Model for Early Diagnosis of Alzheimer's Disease. J Med Syst. 2016; 40:125. [PubMed: 27059738]

Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, Rei L, Squarcia S, Rodriguez G, Bellotti R, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage. 2011; 58:469–480. [PubMed: 21718788]

Cho Y, Seong JK, Jeong Y, Shin SY. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage. 2012; 59:2217–2230. [PubMed: 22008371]

Choo IH, Lee DY, Oh JS, Lee JS, Lee DS, Song IC, Youn JC, Kim SG, Kim KW, Jhoo JH, et al. Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and Alzheimer's disease. Neurobiol Aging. 2010; 31:772–779. [PubMed: 18687503]

Chu C, Hsu AL, Chou KH, Bandettini P, Lin C. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage. 2012; 60:59–70. [PubMed: 22166797]

Chu C, Ni Y, Tan G, Saunders CJ, Ashburner J. Kernel regression for fMRI pattern prediction. Neuroimage. 2011; 56:662–673. [PubMed: 20348000]

Clark DG, McLaughlin PM, Woo E, Hwang K, Hurtz S, Ramirez L, Eastman J, Dukes RM, Kapur P, DeRamus TP, et al. Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment. Alzheimers Dement (Amst). 2016; 2:113–122. [PubMed: 27239542]

Costafreda SG, Dinov ID, Tu Z, Shi Y, Liu CY, Kloszewska I, Mecocci P, Soininen H, Tsolaki M, Vellas B, et al. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. Neuroimage. 2011; 56:212–219. [PubMed: 21272654]

Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS One. 2011; 6:e21896. [PubMed: 21814561]

Cui Y, Wen W, Lipnicki DM, Beg MF, Jin JS, Luo S, Zhu W, Kochan NA, Reppermund S, Zhuang L, et al. Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach. Neuroimage. 2012; 59:1209–1217. [PubMed: 21864688]

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, Chupin M, Benali H, Colliot O. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage. 2011; 56:766–781. [PubMed: 20542124]

Da X, Toledo JB, Zee J, Wolk DA, Xie SX, Ou Y, Shacklett A, Parmpi P, Shaw L, Trojanowski JQ, et al. Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. Neuroimage Clin. 2014; 4:164–173. [PubMed: 24371799]

Dai Z, Yan C, Wang Z, Wang J, Xia M, Li K, He Y. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). Neuroimage. 2012; 59:2187–2195. [PubMed: 22008370]

Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging. 2011; 32:2322, e2319–2327.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging. 2008; 29:514–523. [PubMed: 17174012]

Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage. 2001; 14:1361–1369. [PubMed: 11707092]

Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain. 2009; 132:2026–2035. [PubMed: 19416949]

Dennis EL, Thompson PM. Functional brain connectivity using fMRI in aging and Alzheimer's disease. Neuropsychol Rev. 2014; 24:49–62. [PubMed: 24562737]

Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL, et al. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain. 2009; 132:2048–2057. [PubMed: 19460794]

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006; 31:968–980. [PubMed: 16530430]

Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, Grodstein F, Wright CI, Blacker D, Rosas HD, et al. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb Cortex. 2009; 19:497–510. [PubMed: 18632739]

Dickerson BC, Goncharova I, Sullivan MP, Forchetti C, Wilson RS, Bennett DA, Beckett LA, deToledo-Morrell L. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiol Aging. 2001; 22:747–754. [PubMed: 11705634]

Dickerson BC, Stoub TR, Shah RC, Sperling RA, Killiany RJ, Albert MS, Hyman BT, Blacker D, Detoledo-Morrell L. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. Neurology. 2011; 76:1395–1402. [PubMed: 21490323]

Dong A, Honnorat N, Gaonkar B, Davatzikos C. CHIMERA: Clustering of Heterogeneous Disease Effects via Distribution Matching of Imaging Patterns. IEEE Trans Med Imaging. 2016a; 35:612–621. [PubMed: 26452275]

Dong A, Toledo JB, Honnorat N, Doshi J, Varol E, Sotiras A, Wolk D, Trojanowski JQ, Davatzikos C. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain. 2016b

Dukart J, Mueller K, Barthel H, Villringer A, Sabri O, Schroeter ML. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry Res. 2013; 212:230–236. [PubMed: 23149027]

Dukart J, Mueller K, Horstmann A, Barthel H, Moller HE, Villringer A, Sabri O, Schroeter ML. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. PLoS One. 2011a; 6:e18111. [PubMed: 21448435]

Dukart J, Schroeter ML, Mueller K. Age correction in dementia-matching to a healthy brain. PLoS One. 2011b; 6:e22193. [PubMed: 21829449]

Dyrba M, Barkhof F, Fellgiebel A, Filippi M, Hausner L, Hauenstein K, Kirste T, Teipel SJ. Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data. J Neuroimaging. 2015a; 25:738–747. [PubMed: 25644739]

Dyrba M, Ewers M, Wegrzyn M, Kilimann I, Plant C, Oswald A, Meindl T, Pievani M, Bokde AL, Fellgiebel A, et al. Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data. PLoS One. 2013; 8:e64925. [PubMed: 23741425]

Dyrba M, Grothe M, Kirste T, Teipel SJ. Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. Hum Brain Mapp. 2015b; 36:2118–2131. [PubMed: 25664619]

Eskildsen SF, Coupé P, Fonov V, Collins DL. Detecting Alzheimer's disease b morphological MRI using hippocampal grading and cortical thickness. Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data. 2014:38–47.

Eskildsen SF, Coupe P, Garcia-Lorenzo D, Fonov V, Pruessner JC, Collins DL. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage. 2013; 65:511–521. [PubMed: 23036450]

Falahati F, Westman E, Simmons A. Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. J Alzheimers Dis. 2014; 41:685–708. [PubMed: 24718104]

Fan Y, Batmanghelich N, Clark CM, Davatzikos C. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage. 2008a; 39:1731–1743. [PubMed: 18053747]

Fan Y, Liu Y, Wu H, Hao Y, Liu H, Liu Z, Jiang T. Discriminant analysis of functional connectivity patterns on Grassmann manifold. Neuroimage. 2011; 56:2058–2067. [PubMed: 21440643]

Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage. 2008b; 41:277–285. [PubMed: 18400519]

Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans Med Imaging. 2007; 26:93–105. [PubMed: 17243588]

Fellgiebel A, Dellani PR, Greverus D, Scheurich A, Stoeter P, Muller MJ. Predicting conversion to dementia in mild cognitive impairment by volumetric and diffusivity measurements of the hippocampus. Psychiatry Res. 2006; 146:283–287. [PubMed: 16530394]

Fellgiebel A, Muller MJ, Wille P, Dellani PR, Scheurich A, Schmidt LG, Stoeter P. Color-coded diffusion-tensor-imaging of posterior cingulate fiber tracts in mild cognitive impairment. Neurobiol Aging. 2005; 26:1193–1198. [PubMed: 15917103]

Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000; 97:11050–11055. [PubMed: 10984517]

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 1999; 8:272–284. [PubMed: 10619420]

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 2007; 8:700–711. [PubMed: 17704812]

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A. 2005; 102:9673–9678. [PubMed: 15976020]

Fransson P. Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. Hum Brain Mapp. 2005; 26:15–29. [PubMed: 15852468]

Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007; 315:972–976. [PubMed: 17218491]

Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol. 2010; 6:67–77. [PubMed: 20139996]

Frisoni GB, Testa C, Zorzan A, Sabattoli F, Beltramello A, Soininen H, Laakso MP. Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. J Neurol Neurosurg Psychiatry. 2002; 73:657–664. [PubMed: 12438466]

Gerardin E, Chetelat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, Niethammer M, Dubois B, Lehericy S, Garnero L, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage. 2009; 47:1476–1486. [PubMed: 19463957]

Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. Neuroimage. 2013; 65:167–175. [PubMed: 23041336]

Gray KR, Wolz R, Heckemann RA, Aljabar P, Hammers A, Rueckert D. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage. 2012; 60:221–229. [PubMed: 22236449]

Greicius MD, Krasnow B, Reiss AL, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A. 2003; 100:253–258. [PubMed: 12506194]

Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci U S A. 2004; 101:4637–4642. [PubMed: 15070770]

Habes M, Erus G, Toledo JB, Zhang T, Bryan N, Launer LJ, Rosseel Y, Janowitz D, Doshi J, Van der Auwera S, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. Brain. 2016a; 139:1164–1179. [PubMed: 26912649]

Habes M, Janowitz D, Erus G, Toledo JB, Resnick SM, Doshi J, Van der Auwera S, Wittfeld K, Hegenscheid K, Hosten N, et al. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. Transl Psychiatry. 2016b; 6:e775. [PubMed: 27045845]

Habes M, Toledo JB, Resnick SM, Doshi J, Van der Auwera S, Erus G, Janowitz D, Hegenscheid K, Homuth G, Volzke H, et al. Relationship between APOE Genotype and Structural MRI Measures throughout Adulthood in the Study of Health in Pomerania Population-Based Cohort. AJNR Am J Neuroradiol. 2016c; 37:1636–1642. [PubMed: 27173368]

He Y, Wang L, Zang Y, Tian L, Zhang X, Li K, Jiang T. Regional coherence changes in the early stages of Alzheimer's disease: a combined structural and resting-state functional MRI study. Neuroimage. 2007; 35:488–500. [PubMed: 17254803]

Herholz K, Salmon E, Perani D, Baron JC, Holthoff V, Frolich L, Schonknecht P, Ito K, Mielke R, Kalbe E, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. Neuroimage. 2002; 17:302–316. [PubMed: 12482085]

Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage. 2009; 48:138–149. [PubMed: 19481161]

Hinrichs C, Singh V, Xu G, Johnson SC. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage. 2011; 55:574–589. [PubMed: 21146621]

Ishii K, Kawachi T, Sasaki H, Kono AK, Fukuda T, Kojima Y, Mori E. Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. AJNR Am J Neuroradiol. 2005; 26:333–340. [PubMed: 15709131]

Ithapu VK, Singh V, Okonkwo OC, Chappell RJ, Dowling NM, Johnson SC. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. Alzheimers Dement. 2015; 11:1489–1499. [PubMed: 26093156]

Jagust W. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. Neuron. 2013; 77:219–234. [PubMed: 23352159]

Jie B, Zhang D, Gao W, Wang Q, Wee CY, Shen D. Integration of network topological and connectivity properties for neuroimaging classification. IEEE Trans Biomed Eng. 2014; 61:576–589. [PubMed: 24108708]

Jones SE, Buchbinder BR, Aharon I. Three-dimensional mapping of cortical thickness using Laplace's equation. Hum Brain Mapp. 2000; 11:12–32. [PubMed: 10997850]

Kantarci K, Jack CR Jr, Xu YC, Campeau NG, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Kokmen E, Tangalos EG, et al. Mild cognitive impairment and Alzheimer disease: regional diffusivity of water. Radiology. 2001; 219:101–107. [PubMed: 11274543]

Khazaee A, Ebrahimzadeh A, Babajani-Feremi A. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. Clin Neurophysiol. 2015; 126:2132–2141. [PubMed: 25907414]

Khondoker M, Dobson R, Skirrow C, Simmons A, Stahl D. A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. Stat Methods Med Res. 2016; 25:1804–1823. [PubMed: 24047600]

Killiany RJ, Hyman BT, Gomez-Isla T, Moss MB, Kikinis R, Jolesz F, Tanzi R, Jones K, Albert MS. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. Neurology. 2002; 58:1188–1196. [PubMed: 11971085]

Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RS. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008; 131:681–689. [PubMed: 18202106]

Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, Bergstrom M, Savitcheva I, Huang GF, Estrada S, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol. 2004; 55:306–319. [PubMed: 14991808]

Koch W, Teipel S, Mueller S, Benninghoff J, Wagner M, Bokde AL, Hampel H, Coates U, Reiser M, Meindl T. Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. Neurobiol Aging. 2012; 33:466–478. [PubMed: 20541837]

Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging. 2010; 31:1429–1442. [PubMed: 20541286]

Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Articial Intelligence. 1995:1137–1145.

Koivunen J, Scheinin N, Virta JR, Aalto S, Vahlberg T, Nagren K, Helin S, Parkkola R, Viitanen M, Rinne JO. Amyloid PET imaging in patients with mild cognitive impairment: a 2-year follow-up study. Neurology. 2011; 76:1085–1090. [PubMed: 21325653]

Korolev IO, Symonds LL, Bozoki AC. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. PLoS One. 2016; 11:e0138866. [PubMed: 26901338]

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009; 12:535–540. [PubMed: 19396166]

Langbaum JB, Chen K, Lee W, Reschke C, Bandy D, Fleisher AS, Alexander GE, Foster NL, Weiner MW, Koeppe RA, et al. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Neuroimage. 2009; 45:1107–1116. [PubMed: 19349228]

Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund LO, Dodge Y, Dierks T. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). J Neurosci Methods. 2007; 161:342–350. [PubMed: 17156848]

Lemm S, Blankertz B, Dickhaus T, Muller KR. Introduction to machine learning for brain imaging. Neuroimage. 2011; 56:387–399. [PubMed: 21172442]

Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, Evans AC. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging. 2008; 29:23–30. [PubMed: 17097767]

Li L, Rakitsch B, Borgwardt K. ccSVM: correcting Support Vector Machines for confounding factors in biological data classification. Bioinformatics. 2011; 27:i342–348. [PubMed: 21685091]

Li M, Qin Y, Gao F, Zhu W, He X. Discriminative analysis of multivariate features from structural MRI and diffusion tensor images. Magn Reson Imaging. 2014a; 32:1043–1051. [PubMed: 24970026]

Li S, Shi F, Pu F, Li X, Jiang T, Xie S, Wang Y. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. AJNR Am J Neuroradiol. 2007; 28:1339–1345. [PubMed: 17698538]

Li S, Yuan X, Pu F, Li D, Fan Y, Wu L, Chao W, Chen N, He Y, Han Y. Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients. J Neurosci. 2014b; 34:10541–10553. [PubMed: 25100588]

Li SJ, Li Z, Wu G, Zhang MJ, Franczak M, Antuono PG. Alzheimer Disease: evaluation of a functional MR imaging index as a marker. Radiology. 2002; 225:253–259. [PubMed: 12355013]

Lillemark L, Sorensen L, Pai A, Dam EB, Nielsen M. Brain region's relative proximity as marker for Alzheimer's disease based on structural MRI. BMC Med Imaging. 2014; 14:21. [PubMed: 24889999]

Liu F, Zhou L, Shen C, Yin J. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. IEEE J Biomed Health Inform. 2014; 18:984–990. [PubMed: 24132030]

Liu M, Zhang D, Shen D. View-centralized multi-atlas classification for Alzheimer's disease diagnosis. Hum Brain Mapp. 2015; 36:1847–1865. [PubMed: 25624081]

Liu X, Tosun D, Weiner MW, Schuff N. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. Neuroimage. 2013; 83:148–157. [PubMed: 23792982]

Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund LO, et al. AddNeuroMed–the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. Ann N Y Acad Sci. 2009; 1180:36–46. [PubMed: 19906259]

Lovestone S, Francis P, Strandgaard K. Biomarkers for disease modification trials–the innovative medicines initiative and AddNeuroMed. J Nutr Health Aging. 2007; 11:359–361. [PubMed: 17653500]

MacDonald D, Kabani N, Avis D, Evans AC. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. Neuroimage. 2000; 12:340–356. [PubMed: 10944416]

Madsen SK, Ho AJ, Hua X, Saharan PS, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. Neurobiol Aging. 1312; 31:1312–1325.

Magnin B, Mesrob L, Kinkingnehun S, Pelegrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehericy S, Benali H. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology. 2009; 51:73–83. [PubMed: 18846369]

McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ Jr, Holland D, Karow DS, Pung CJ, Brewer JB, Dale AM. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology. 2009; 251:195–205. [PubMed: 19201945]

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al. 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. 2011; 7:263–269. [PubMed: 21514250]

Medina D, DeToledo-Morrell L, Urresta F, Gabrieli JD, Moseley M, Fleischman D, Bennett DA, Leurgans S, Turner DA, Stebbins GT. White matter changes in mild cognitive impairment and AD: A diffusion tensor imaging study. Neurobiol Aging. 2006; 27:663–672. [PubMed: 16005548]

Melah KE, Lu SY, Hoscheidt SM, Alexander AL, Adluru N, Destiche DJ, Carlsson CM, Zetterberg H, Blennow K, Okonkwo OC, et al. Cerebrospinal Fluid Markers of Alzheimer's Disease Pathology and Microglial Activation are Associated with Altered White Matter Microstructure in Asymptomatic Adults at Risk for Alzheimer's Disease. J Alzheimers Dis. 2016; 50:873–886. [PubMed: 26836182]

Mendelson AF, Zuluaga MA, Thurfjell L, Hutton BF, Ourselin S. The empirical variance estimator for computer aided diagnosis: lessons for algorithm validation. International Conference on Medical Image Computing and Computer-Assisted Intervention. 2014:236–243.

Min R, Wu G, Cheng J, Wang Q, Shen D. Multi-atlas based representations for Alzheimer's disease diagnosis. Hum Brain Mapp. 2014; 35:5052–5070. [PubMed: 24753060]

Misra C, Fan Y, Davatzikos C. 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. 2009; 44:1415–1422. [PubMed: 19027862]

Moller C, Pijnenburg YA, van der Flier WM, Versteeg A, Tijms B, de Munck JC, Hafkemeijer A, Rombouts SA, van der Grond J, van Swieten J, et al. Alzheimer Disease and Behavioral Variant Frontotemporal Dementia: Automatic Classification Based on Cortical Atrophy for Single-Subject Diagnosis. Radiology. 2016; 279:838–848. [PubMed: 26653846]

Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage. 2015; 104:398–412. [PubMed: 25312773]

Mosconi L, Tsui WH, De Santi S, Li J, Rusinek H, Convit A, Li Y, Boppana M, de Leon MJ. Reduced hippocampal metabolism in MCI and AD: automated FDG-PET image analysis. Neurology. 2005; 64:1860–1867. [PubMed: 15955934]

Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, Reiman EM, Holthoff V, Kalbe E, Sorbi S, et al. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. J Nucl Med. 2008; 49:390–398. [PubMed: 18287270]

Muller MJ, Greverus D, Dellani PR, Weibrich C, Wille PR, Scheurich A, Stoeter P, Fellgiebel A. Functional implications of hippocampal volume and diffusivity in mild cognitive impairment. Neuroimage. 2005; 28:1033–1042. [PubMed: 16084115]

Muller MJ, Greverus D, Weibrich C, Dellani PR, Scheurich A, Stoeter P, Fellgiebel A. Diagnostic utility of hippocampal size and mean diffusivity in amnestic MCI. Neurobiol Aging. 2007; 28:398–403. [PubMed: 16529847]

Murp M, LeVine H III. Alzheimer's disease and the β-amyloid peptide. J Alzheimers Dis. 2010; 19:311–323. [PubMed: 20061647]

Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 2011; 10:785–796. [PubMed: 21802369]

Nettiksimmons J, DeCarli C, Landau S, Beckett L. Biological heterogeneity in ADNI amnestic mild cognitive impairment. Alzheimers Dement. 2014; 10:511–521. e511. [PubMed: 24418061]

Nir TM, Villalon-Reina JE, Prasad G, Jahanshad N, Joshi SH, Toga AW, Bernstein MA, Jack CR Jr, Weiner MW, Thompson PM. Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease. Neurobiol Aging. 2015; 36(Suppl 1):S132–140. [PubMed: 25444597]

Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, Ye BS, Yoon CW, Kim HJ, Chin J, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology. 2014; 83:1936–1944. [PubMed: 25344382]

Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, Luxen A, Phillips C, Laureys S. Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. Neuroimage Clin. 2014; 4:687–694. [PubMed: 24936420]

Okello A, Koivunen J, Edison P, Archer HA, Turkheimer FE, Nagren K, Bullock R, Walker Z, Kennedy A, Fox NC, et al. Conversion of amyloid positive and negative MCI to AD over 3 years: an 11C-PIB PET study. Neurology. 2009; 73:754–760. [PubMed: 19587325]

Oliveira PP Jr, Nitrini R, Busatto G, Buchpiguel C, Sato JR, Amaro E Jr. Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease. J Alzheimers Dis. 2010; 19:1263–1272. [PubMed: 20061613]

Ota K, Oishi N, Ito K, Fukuyama H. A comparison of three brain atlases for MCI prediction. J Neurosci Methods. 2014; 221:139–150. [PubMed: 24140118]

Ota K, Oishi N, Ito K, Fukuyama H. Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease. J Neurosci Methods. 2015; 256:168–183. [PubMed: 26318777]

Padilla P, Lopez M, Gorriz JM, Ramirez J, Salas-Gonzalez D, Alvarez I. NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease. IEEE Trans Med Imaging. 2012; 31:207–216. [PubMed: 21914569]

Pagani M, De Carli F, Morbelli S, Oberg J, Chincarini A, Frisoni GB, Galluzzi S, Perneczky R, Drzezga A, van Berckel BN, et al. Volume of interest-based [18F]fluorodeoxyglucose PET discriminates MCI converting to Alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study. Neuroimage Clin. 2015; 7:34–42. [PubMed: 25610765]

Park H, Yang JJ, Seo J, Lee JM. Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment. Neurosci Lett. 2012; 529:123–127. [PubMed: 23000551]

Park H, Yang JJ, Seo J, Lee JM. Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease. Neurosci Lett. 2013; 550:17–22. [PubMed: 23827219]

Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005; 27:1226–1238. [PubMed: 16119262]

Perrin RJ, Fagan AM, Holtzman DM. Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. Nature. 2009; 461:916–922. [PubMed: 19829371]

Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med. 2004; 256:183–194. [PubMed: 15324362]

Pievani M, de Haan W, Wu T, Seeley WW, Frisoni GB. Functional network disruption in the degenerative dementias. Lancet Neurol. 2011; 10:829–843. [PubMed: 21778116]

Plant C, Teipel SJ, Oswald A, Bohm C, Meindl T, Mourao-Miranda J, Bokde AW, Hampel H, Ewers M. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage. 2010; 50:162–174. [PubMed: 19961938]

Prasad G, Joshi SH, Nir TM, Toga AW, Thompson PM. Brain connectivity and novel network measures for Alzheimer's disease classification. Neurobiol Aging. 2015; 36(Suppl 1):S121–131. [PubMed: 25264345]

Querbes O, Aubry F, Pariente J, Lotterie JA, Demonet JF, Duret V, Puel M, Berry I, Fort JC, Celsis P. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain. 2009; 132:2036–2047. [PubMed: 19439419]

Ridha BH, Anderson VM, Barnes J, Boyes RG, Price SL, Rossor MN, Whitwell JL, Jenkins L, Black RS, Grundman M, et al. Volumetric MRI and cognitive measures in Alzheimer disease : comparison of markers of progression. J Neurol. 2008; 255:567–574. [PubMed: 18274807]

Rose SE, McMahon KL, Janke AL, O'Dowd B, de Zubicaray G, Strudwick MW, Chalk JB. Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnestic mild cognitive impairment. J Neurol Neurosurg Psychiatry. 2006; 77:1122–1128. [PubMed: 16754694]

Sabuncu MR, Konukoglu E. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinformatics. 2015; 13:31–46. [PubMed: 25048627]

Sajda P. Machine learning for detection and diagnosis of disease. Annu Rev Biomed Eng. 2006; 8:537–565. [PubMed: 16834566]

Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front Neurosci. 2015; 9:307. [PubMed: 26388719]

Schouten TM, Koini M, de Vos F, Seiler S, van der Grond J, Lechner A, Hafkemeijer A, Moller C, Schmidt R, de Rooij M, et al. Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. Neuroimage Clin. 2016; 11:46–51. [PubMed: 26909327]

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci. 2007; 27:2349–2356. [PubMed: 17329432]

Selnes P, Aarsland D, Bjornerud A, Gjerstad L, Wallin A, Hessen E, Reinvang I, Grambaite R, Auning E, Kjaervik VK, et al. Diffusion tensor imaging surpasses cerebrospinal fluid as predictor of cognitive decline and medial temporal lobe atrophy in subjective cognitive impairment and mild cognitive impairment. J Alzheimers Dis. 2013; 33:723–736. [PubMed: 23186987]

Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological alterations in Alzheimer disease. Cold Spring Harb Perspect Med. 1101; 1:a006189.

Shao Y, Lunetta RS. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing. 2012; 70:78–87.

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage. 2008; 39:1064–1080. [PubMed: 18037310]

Shen KK, Fripp J, Meriaudeau F, Chetelat G, Salvado O, Bourgeat P. Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models. Neuroimage. 2012; 59:2155–2166. [PubMed: 22037419]

Sorensen L, Igel C, Liv Hansen N, Osler M, Lauritzen M, Rostrup E, Nielsen M. Early detection of Alzheimer's disease using MRI hippocampal texture. Hum Brain Mapp. 2016; 37:1148–1161. [PubMed: 26686837]

Sørensen L, Pai A, Anker C, Balas I, Lillholm M, Igel C, Nielsen M. Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry. Proc MICCAI workshop challenge on computer-aided diagnosis of dementia based on structural MRI data. 2014:111–118.

Tang X, Holland D, Dale AM, Younes L, Miller MI. Baseline shape diffeomorphometry patterns of subcortical and ventricular structures in predicting conversion of mild cognitive impairment to Alzheimer's disease. J Alzheimers Dis. 2015; 44:599–611. [PubMed: 25318546]

Tang X, Qin Y, Wu J, Zhang M, Zhu W, Miller MI. Shape and diffusion tensor imaging based integrative analysis of the hippocampus and the amygdala in Alzheimer's disease. Magn Reson Imaging. 2016; 34:1087–1099. [PubMed: 27211255]

Teipel S, Drzezga A, Grothe MJ, Barthel H, Chetelat G, Schuff N, Skudlarski P, Cavedo E, Frisoni GB, Hoffmann W, et al. Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection. Lancet Neurol. 2015; 14:1037–1053. [PubMed: 26318837]

Termenon M, Graña M. A two stage sequential ensemble applied to the classification of Alzheimer's disease based on mri features. Neural rocessing Letters. 2012; 35:1–12.

Tipping ME. Sparse Bayesian learning and the relevance vector machine. Journal of machine learning research. 2001; 1:211–244.

Toussaint PJ, Perlbarg V, Bellec P, Desarnaud S, Lacomblez L, Doyon J, Habert MO, Benali H. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. Neuroimage. 2012; 63:936–946. [PubMed: 22510256]

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002; 15:273–289. [PubMed: 11771995]

Vandenberghe R, Nelissen N, Salmon E, Ivanoiu A, Hasselbalch S, Andersen A, Korner A, Minthon L, Brooks DJ, Van Laere K, et al. Binary classification of (1)(8)F-flutemetamol PET using machine learning: comparison with visual reads and structural MRI. Neuroimage. 2013; 64:517–525. [PubMed: 22982358]

Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw. 1999; 10:988–999. [PubMed: 18252602]

Varol E, Sotiras A, Davatzikos C. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage. 2017; 145:346–364. [PubMed: 26923371]

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage. 2008; 39:1186–1197. [PubMed: 18054253]

Vemuri P, Jack CR Jr. Role of structural MRI in Alzheimer's disease. Alzheimers Res Ther. 2010; 2:23. [PubMed: 20807454]

Visser PJ, Scheltens P, Verhey FR, Schmand B, Launer LJ, Jolles J, Jonker C. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. J Neurol. 1999; 246:477–485. [PubMed: 10431775]

Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T. Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. Hum Brain Mapp. 2007a; 28:967–978. [PubMed: 17133390]

Wang L, Beg F, Ratnanather T, Ceritoglu C, Younes L, Morris JC, Csernansky JG, Miller MI. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. IEEE Trans Med Imaging. 2007b; 26:462–470. [PubMed: 17427733]

Wang L, Zang Y, He Y, Liang M, Zhang X, Tian L, Wu T, Jiang T, Li K. Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. Neuroimage. 2006; 31:496–504. [PubMed: 16473024]

Wee CY, Yap PT, Li W, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Enriched white matter connectivity networks for accurate identification of MCI patients. Neuroimage. 2011; 54:1812–1822. [PubMed: 20970508]

Wee CY, Yap PT, Shen D. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum Brain Mapp. 2013; 34:3411–3425. [PubMed: 22927119]

Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Cedarbaum J, Green RC, Harvey D, Jack CR, Jagust W, et al. 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. Alzheimers Dement. 2015; 11:e1–120. [PubMed: 26073027]

Westman E, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Weiner MW, Lovestone S, et al. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage. 2011; 58:818–828. [PubMed: 21763442]

Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, Knopman DS, Boeve BF, Parisi JE, Petersen RC, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. Lancet Neurol. 2012; 11:868–877. [PubMed: 22951070]

Wisse LE, Biessels GJ, Heringa SM, Kuijf HJ, Koek DH, Luijten PR, Geerlings MI. Hippocampal subfield volumes at 7T in early Alzheimer's disease and normal aging. Neurobiol Aging. 2014; 35:2039–2045. [PubMed: 24684788]

Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, Soininen H, Lotjonen J. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS One. 2011; 6:e25446. [PubMed: 22022397]

Xie S, Xiao JX, Gong GL, Zang YF, Wang YH, Wu HK, Jiang XX. Voxel-based detection of white matter abnormalities in mild Alzheimer disease. Neurology. 2006; 66:1845–1849. [PubMed: 16801648]

Xu L, Wu X, Chen K, Yao L. Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. Comput Methods Programs Biomed. 2015; 122:182–190. [PubMed: 26298855]

Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. Neuroimage Clin. 2013; 2:735–745. [PubMed: 24179825]

Yu G, Liu Y, Shen D. Graph-guided joint prediction of class label and clinical scores for the Alzheimer's disease. Brain Struct Funct. 2016; 221:3787–3801. [PubMed: 26476928]

Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage. 2012; 59:895–907. [PubMed: 21992749]

Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011; 55:856–867. [PubMed: 21236349]

Zhang Y, Wang S, Dong Z. Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. Progress In Electromagnetics Research. 2014; 144:171–184.

Zheng W, Yao Z, Hu B, Gao X, Cai H, Moore P. Novel Cortical Thickness Pattern for Accurate Detection of Alzheimer's Disease. J Alzheimers Dis. 2015; 48:995–1008. [PubMed: 26444768]

Zhu X, Suk HI, Shen D. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. Neuroimage. 2014; 100:91–105. [PubMed: 24911377]

Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005; 67:301–320.

Zu C, Jie B, Liu M, Chen S, Shen D, Zhang D. Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. Brain Imaging Behav. 2015; 10:1148–1159.

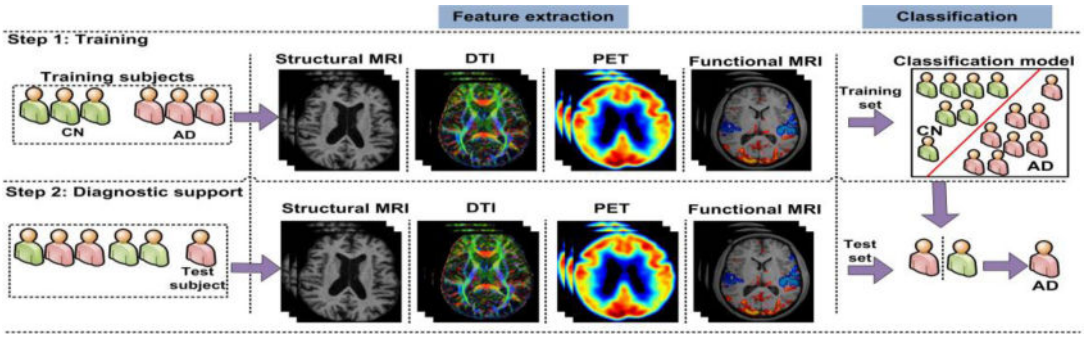

Figure 1.
A top-level layout of neuroimaging-based classification framework for AD classification

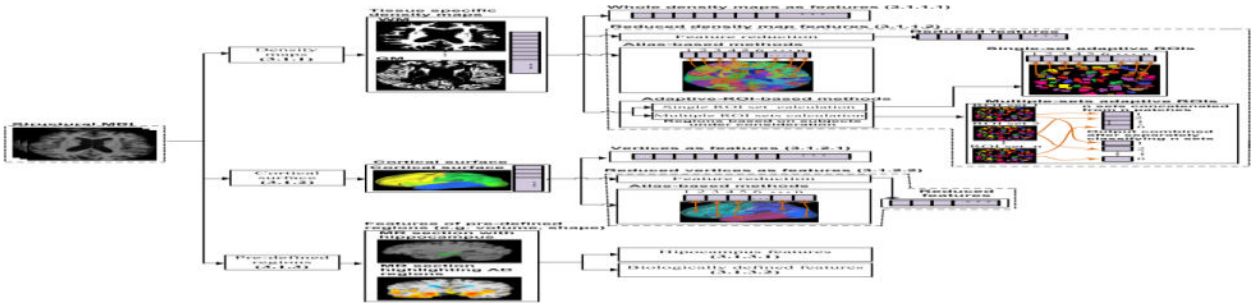

Figure 2.
A top-level breakdown of structural MRI-based classification studies for AD classification

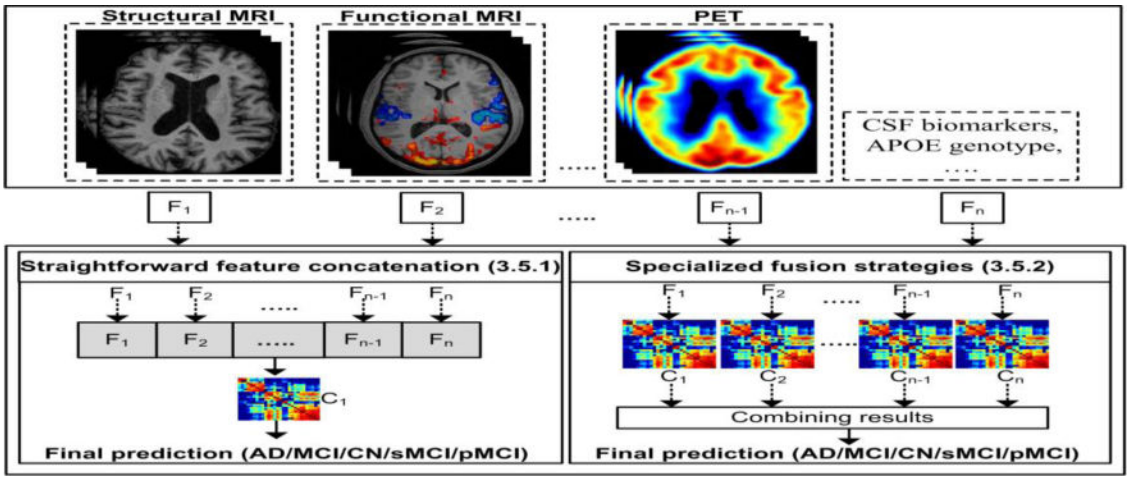

Figure 3.

A top-level breakdown of multimodality-based classification studies for AD classification

Table 1

A brief description of the datasets used for the validation of density map-based AD classification frameworks

Study	Subjects					Type	Classification algorithm	Database	Classification accuracy		
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	sMCI/pMCI
(Kloppel et al., 2008)	20	—	—	—	20	DMAF	SVM	RM	95.00	—	—
(Kloppel et al., 2008)	14	—	—	—	14	DMAF	SVM	DRCUCL	92.90	—	—
(Kloppel et al., 2008)	33	—	—	—	57	DMAF	SVM	RM	81.10	—	—
(Casanova et al., 2011)	49	—	—	—	49	DMAF	LSR	ADNI	85.70	—	—
(Hinrichs et al., 2009)	89	—	—	—	94	DMAF	LPBM	ADNI	82.00	—	—
(Termenon, Graña, 2012)	49	—	—	—	49	DMAF	RVM,SVM	OASIS	83.00	—	—
(Adaszewski et al., 2013)	108	—	142	61	137	DMAF	SVM	ADNI	—	—	62.00
(Plant et al., 2010)	32	—	9	15	18	DMAF	SVM,BS,VFI	CPLMU	92.00	—	75.00
(Moller et al., 2016)	84	—	—	—	94	DMAF	SVM	ACVUMC+ACEUMC	88.00	—	—
(Liu et al., 2013)	86	—	97	93	137	SUFR	RLR,SVM,LDA	ADNI	90.00	—	68.00
(Salvatore et al., 2015)	137	—	76	134	162	SUFR	SVM	ADNI	76.00	72.00[1]	66.00
(Beheshti, Demirel, 2015)	130	—	—	—	130	SUFR	SVM	ADNI	89.65	—	—
(Magnin et al., 2009)	16	–	—	—	22	Atlas	SVM	Internal	94.50	—	—
(Fan et al., 2008a)	56	88	—	—	66	S-ROI	SVM	ADNI	94.30	—	—
(Davatzikos et al., 2008)	—	15	—	—	15	S-ROI	SVM	BLSA	—	90.00	—
(Misra et al., 2009)	56	—	27	76	66	S-ROI	SVM	ADNI	—	—	81.50
(Min et al., 2014)	97	—	117	117	128	M-ROI	SVM	ADNI	91.64	—	72.41
(Liu et al., 2015)	97	—	117	117	128	M-ROI	SVM	ADNI	92.51	–	78.88

SUFR = Supervised/unsupervised feature reduction

S(M)-ROI = Single(Multiple)-set adaptive ROIs

LSR = Large scale regularization

VFI = Voting feature intervals

BS = Bayes statistics

OASIS = Open access series of imaging initiative (Chan et al., 2003)

RM = Rochester, Minnesota, USA

DRCUCL = Dementia research center, University College London

CPLMU = Clinic of psychiatry at the Ludwig Maximilian university, of Munich

ACVUMC = Alzheimer center of the VU university medical center

ACEUMC = Alzheimer center of the Erasmus university medical center

1 = cMCI/CN

In all the tables, blank entry (–) means information was not provided by the authors or is irrelevant.

Table 2

A brief description of the datasets used for the validation of cortical thickness-based AD classification frameworks

Study	Subjects					Type	Classification algorithm	Database	Classification accuracy			
	AD	MCI	pMCI	sMCI	CN				AD/CN	AD/MCI	MCI/CN	sMCI/pMCI
(Li et al., 2014b)	—	24	—	—	26	VAF	SVM	XWH	—	—	80.00	—
(Cho et al., 2012)	128	—	72	131	160	RV	LDA	ADNI	88.33	—	—	71.21
(Park et al., 2012)	25	25	—	—	50	RV	SVM	OASIS	90.00	90.00	86.00	—
(Park et al., 2013)	—	30	12	—	30	RV	SVM	ADNI	—	83.00[1]	90.00	—
(Desikan et al., 2009)	65	57	—	—	94	Atlas	LR	ADNI	95.00	—	95.00	—
(McEvoy et al., 2009)	84	175	—	—	139	Atlas	LDA	ADNI	89.00	—	—	—
(Oliveira et al., 2010)	14	—	—	—	20	Atlas	SVM	DPUSP	88.20	—	—	—
(Eskildsen et al., 2013)	194	—	340	134	226	Atlas	LDA	ADNI	86.70	—	—	71.10
(Eskildsen et al., 2013)	—	—	29 (36)	134	–	Atlas	LDA	ADNI	—	—	—	77.30
(Eskildsen et al., 2013)	—	—	61 (24)	134	—	Atlas	LDA	ADNI	—	—	—	73.00
(Eskildsen et al., 2013)	—	—	128 (12)	134	—	Atlas	LDA	ADNI	—	—	—	74.50
(Eskildsen et al., 2013)	—	—	122 (06)	134	—	Atlas	LDA	ADNI	—	—	—	80.90
(Wee et al., 2013)	198	—	89	111	200	Atlas	Multi-kernel SVM	ADNI	92.35	79.24	83.75	75.05
(Lillemark et al., 2014)	114	240	—	—	170	Atlas	LDA	ADNI	87.70*	76.60*	78.40*	—

RV = Reduced vertex-wise

XWH = Xuan Wu Hospital, Beijing, China

D Department of s chiatr, niversit of ao aulo, ao aulo, razil

[1] = prediction of conversion from MCI to AD

* shows AUC.

The first entry for (Eskildsen et al., 2013) shows the classification accuracy by using all the sMCIs and pMCIs. The remaining entries show the classification between sMCI and different sets of pMCI subjects, divided according to their time of conversion from sMCI to pMCI. Numbers of months are shown in parenthesis against pMCI subjects.

Table 3

A brief description of the datasets used for the validation of pre-defined regions-based AD classification frameworks

Study	Subjects					Type	Classification algorithm	Database	Classification accuracy		
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	sMCI/pMCI
(Wang et al., 2007b)	18	—	—	—	26	Hippocampus	LR	ADRC	81.10	—	—
(Li et al., 2007)	19	—	—	—	20	Hippocampus	SVM	Internal	94.90	—	—
(Gerardin et al., 2009)	23	23	—	—	25	Hippocampus	SVM	HCUDC	94.00	83.00	—
(Shen et al., 2012)	99	—	—	—	138	Hippocampus	Bagged SVM	ADNI	88.30	—	—
(Costafreda et al., 2011)	71	—	22	81	88	Hippocampus	SVM	AddNeuroMed	—	—	80.00
(Sorensen et al., 2016)	101	233	93	140	169	Hippocampus	SVM	ADNI	91.20*	76.40*	74.20*
(Chincarini et al., 2011)	144	—	136	166	189	BSR	SVM	ADNI	97.00*	92.00[1,*]	74.00*
(Tang et al., 2015)	175	—	135	87	210	BSR	LDA	ADNI	—	—	74.77

BSR = Biologically selected regions

ADRC = Alzheimer disease research center, Washington University School of Medicine, St. Louis, Missouri.

HCUDC = Hospital center University De Caen, Caen, France.

[1] = cMCI versus CN

* shows AUC.

Table 4

A brief description of the datasets used for the validation of functional MRI based AD classification frameworks

Study	Subjects			Type	Classification algorithm	Database	Classification accuracy		
	AD	MCI	CN				AD/CN	MCI/CN	AD/MCI
(Chen et al., 2011)	20	15	20	Connectivity	Fisher LDA	MDC	82.00[1]	91.00	
(Challis et al., 2015)	27	50	39	Connectivity	GP-LR	Internal	—	75.00	97. 00
(Jie et al., 2014)	—	12	25	Graph measures	Multi-kernel SVM	BIAC	—	91.90	
(Khazaee et al., 2015)	20	—	20	Graph measures	SVM	ADNI	100.00	—	—

GP-LR = Gaussian-process logistic regression

MDC = Memory disorders clinic, Medical College of Wisconsin

BIAC = Brain imaging and analysis center, Duke university, North Carolina, USA

[1] = AD/(MCI+CN)

Table 5

A brief description of the datasets used for the validation of DTI based AD classification frameworks

Study	Subjects					Type	Classification algorithm	Database	Accuracy		
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	sMCI/pMCI
(Nir et al., 2015)	37	113	—	—	50	Tractography	SVM	ADNI	80.60	68.30	—
(Wee et al., 2011)	—	10	—	—	17	Network	SVM	BIAC	—	88.90	—
(Prasad et al., 2015)	38	—	38	74	50	Network	SVM	ADNI	78.20	62.80[1], 59.20[2]	63.40
(Dyrba et al., 2013)	137	—	—	—	143	DVS	SVM	EDSD	83.00	—	—
(Dyrba et al., 2015a)	—	—	35	35	25	DVS	SVM	EDSD	—	77.00	68.00

DVS = Discriminative voxel selection

EDSD = European DTI study on Dementia

BIAC = Brain imaging and analysis center, Duke University, North Carolina, USA

For (Dyrba et al., 2015a), pMCI = amyloid+, sMCI = amyloid-

For (Prasad et al., 2015), pMCI = late MCI, sMCI = early MCI

[1] = CN/late MCIs

[2] = CN/early MCIs

Table 6

A brief description of the datasets used for the validation of FDG-PET and amyloid-PET based AD classification frameworks

Study	Subjects					Type	Classification algorithm	Database	Classification accuracy			
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	AD/MCI	sMCI/pMCI
(Hinrichs et al., 2009)	89	–	–	–	94	VAF	LPBM	ADNI	84.00	—	—	—
(Vandenberghe et al., 2013)	27	20	—	—	25	VAF	SVM	Internal	88.46	—	—	100.00
(Cabral et al., 2015)	—	—	44	56	—	DVS	SVM+GNB	ADNI	—	—	—	85.00
(Cabral et al., 2015)	—	—	26 (6)	56	—	DVS	SVM+GNB	ADNI	—	—	—	86.00
(Cabral et al., 2015)	—	—	41 (12)	56	—	DVS	SVM+GNB	ADNI	—	—	—	82.00
(Cabral et al., 2015)	—	—	33 (18)	56	—	DVS	SVM+GNB	ADNI	—		—	75.00
(Cabral et al., 2015)	—	—	25 (24)	56	—	DVS	SVM+GNB	ADNI	—		—	74.00
(Gray et al., 2012)	50	—	53	64	54	Atlas	SVM	ADNI	88.00	81.30[2]	83.50[1]	63.10
(Pagani et al., 2015)	–	—	62	–	109	Atlas	SVM	EADC	—	91.00[2]	—	—
(Padilla et al., 2012)	53	114	—	—	52	PRO	SVM	ADNI	86.59	—	—	—

PRO = Projection based techniques

DVS = Discriminative voxel selection

EADC = European Alzheimer's disease consortium GNB = Gaussian naive bayes

[1] = AD/sMCI

[1] = CN/pMCI

The first entry for (Cabral et al., 2015) shows the classification accuracy by using all the sMCIs and pMCIs. The remaining entries show the classification between sMCI and different sets of pMCI subjects, divided according to their time of conversion from sMCI to pMCI. Numbers of months are shown in parenthesis against pMCI subjects.

Table 7

A brief description of the datasets used for the validation of various multimodality based AD classification frameworks

	Subjects					Type	Classification algorithm	Database	Classification accuracy			
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	AD/MCI	sMCI/pMCI
(Fan et al., 2008b)	—	15	—	—	15	SFC	SVM	BLSA	—	90.00	—	—
(Vemuri et al., 2008)	190	—	—	—	190	SFC	SVM	ADNI	89.30	—	—	—
(Kohannim et al., 2010)	158	264	—	—	213	SFC	SVM	ADNI	93.81	75.49	—	—
(Davatzikos et al., 2011)	—	—	69	170	—	SFC	SVM	ADNI	—	—	—	61.70
(Dukart et al., 2011a)	21	—	—	—	13	SFC	SVM	Leipzig	100.00	—	—	—
(Cui et al., 2011)	96	—	56	87	111	SFC	SVM	ADNI	—	—	—	67.13
(Cui et al., 2012)	—	79	—	—	204	SFC	SVM	SMAS	—	71.09	—	—
(Dukart et al., 2013)	49	—	—	—	41	SFC	SVM	ANDI+Leipzig	90.00[1]	—	—	—
(Zhang et al., 2014)	24	57	—	—	97	SFC	Kernel SVM decision-tree	OASIS	96.00	85.00	88.00	—
(Zhu et al., 2014)	51	99	—	—	52	SFC	SVM	ADNI	95.90	82.00	—	—
(Li et al., 2014a)	21	—	—	—	15	SFC	SVM	TH	94.30	—	—	—
(Apostolova et al., 2014)	95	182	—	—	111	SFC	SVM	ADNI	85.00	79.00	70.00	—
(Moradi et al., 2015)	200	—	164	100	231	SFC	LDS,Random forest	ADNI	—	—	—	81.72
(Zheng et al., 2015)	163	—	104	94	189	SFC	SVM	ADNI	92.11	—	—	79.37
(Tang et al., 2016)	29	—	—	—	23	SFC	LDA+SVM	TH	94.60	—	—	—
(Schouten et al., 2016)	77	—	—	—	173	SFC	Elastic net classifier	PRODEM	93.00	—	—	—
(Clark et al., 2016)	—	24	—	—	—	SFS	Ensemble classifier	ADRC	—	—	87.20[2]	—
(Hinrichs et al., 2011)	48	119	—	—	66	SFS	Multi-kernel SVM	ADNI	92.40	—	—	—
(Zhang et al., 2011)	51	99	—	—	52	SFS	Multi-kernel SVM	ADNI	93.20	76.40	—	—
(Dai et al., 2012)	16	—	—	—	22	SFS	Ensemble of MU-LDA	XWH	89.47	—	—	—
(Zhang, Shen, 2012)	45	91	—	—	50	SFS	Multi-kernel SVM	ADNI	93.33	83.20	—	—
(Young et al., 2013)	63	—	47	96	73	SFS	Gaussian process classifier	ADNI	—	—	—	74.00
(Gray et al., 2013)	37	—	34	41	35	SFS	Random forest	ADNI	89.00	74.60	—	58.00
(Casanova et al., 2013)	171	—	153	182	188	SFS	RLR	ADNI	87.10	—	—	63.00
(Liu et al., 2014)	50	—	—	—	70	SFS	Multi-kernel SVM	ADNI	87.12	—	—	—
(Xu et al., 2015)	113	—	27	83	117	SFS	SRC	ADNI	94.80	74.50	—	77.80
(Zu et al., 2015)	51	—	43	56	52	SFS	Multi-kernel SVM	ADNI	95.95	80.26	—	69.78

	Subjects					Type	Classification algorithm	Database	Classification accuracy			
	AD	MCI	pMCI	sMCI	CN				AD/CN	MCI/CN	AD/MCI	sMCI/pMCI
(Cheng et al., 2015b)	51	—	43	56	52	SFS	Domain transfer SVM	ADNI	—	86.40	82.70	79.40
(Cheng et al., 2015a)	51	—	43	56	52	SFS	M2TL	ADNI	—	—	—	80.10
(Dyrba et al., 2015b)	28	—	—	—	25	SFS	Multi-kernel SVM	EDSD	85.00	—	—	—
(Korolev et al., 2016)	—	—	139	120	—	SFS	Probabilistic multi-kernel	ADNI	—	—	—	80.00
(Yu et al., 2016)	50	97	—	—	52	SFS	Multi-task learning	ADNI	92.60	80.00	—	—

SFC = Straightforward feature concatenation

SFS = Specialized fusion strategies LDS = Low density separation

SRC = Sparse representation-based classification

MU-LDA = Maximum uncertainty-linear discriminant analysis

M2TL = Multimodal manifold-regularized transfer learning

TH = Tongji hospital, Wuhan, China

SMAS = Sydney memory and aging study

XWH = Xuan wu hospital, Beijing, China

PRODEM = Prospective registry on dementia in Austria

ADRC = Alzheimer disease research center, Washington University school of medicine, St. Louis, Missouri

[1] = Accuracy for the combined ADNI+Leipzig cohort

[2] = Prediction of conversion from MCI to AD (AUC)

Neuroimage. Author manuscript; available in PMC 2018 July 15.