NeuroImage 65 (2013) 167–175

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Random forest-based similarity measures for multi-modal classification of Alzheimer's disease

Katherine R. Gray [a,*], Paul Aljabar [a,b], Rolf A. Heckemann [c,d], Alexander Hammers [c,d], Daniel Rueckert [a] for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK
[b] Biomedical Engineering Department, Division of Imaging Sciences, King's College London, UK
[c] Fondation Neurodis, CERMEP-Imagerie du Vivant, Lyon, France
[d] Centre for Neuroscience (Hammersmith), Faculty of Medicine, Imperial College London, UK

ARTICLE INFO

Article history:
Accepted 24 September 2012
Available online 4 October 2012

Keywords:
Alzheimer's disease
Mild cognitive impairment
Multi-modality classification
Magnetic resonance imaging
Positron emission tomography
Cerebrospinal fluid biomarkers
Genetics

ABSTRACT

Neurodegenerative disorders, such as Alzheimer's disease, are associated with changes in multiple neuroimaging and biological measures. These may provide complementary information for diagnosis and prognosis. We present a multi-modality classification framework in which manifolds are constructed based on pairwise similarity measures derived from random forest classifiers. Similarities from multiple modalities are combined to generate an embedding that simultaneously encodes information about all the available features. Multi-modality classification is then performed using coordinates from this joint embedding. We evaluate the proposed framework by application to neuroimaging and biological data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Features include regional MRI volumes, voxel-based FDG-PET signal intensities, CSF biomarker measures, and categorical genetic information. Classification based on the joint embedding constructed using information from all four modalities out-performs the classification based on any individual modality for comparisons between Alzheimer's disease patients and healthy controls, as well as between mild cognitive impairment patients and healthy controls. Based on the joint embedding, we achieve classification accuracies of 89% between Alzheimer's disease patients and healthy controls, and 75% between mild cognitive impairment patients and healthy controls. These results are comparable with those reported in other recent studies using multi-kernel learning. Random forests provide consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. We demonstrate this by application to data in which the number of features differs by several orders of magnitude between modalities. Random forest classifiers extend naturally to multi-class problems, and the framework described here could be applied to distinguish between multiple patient groups in the future.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Changes in multiple biomarkers may provide complementary information for the diagnosis and prognosis of neurodegenerative disorders such as Alzheimer's disease (AD). At present, the clinical diagnosis of AD is based on assessments of cognition and behaviour, which start to decline in the later disease stages. Recently published revisions to the diagnostic criteria (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011) incorporate suggestions that biological and neuroimaging measures of structural and molecular changes in the brain may be better suited for the early detection of AD, as well as for monitoring its progression. Automated classification of individual patients based on multiple biomarkers could provide valuable support for clinicians, when considered alongside cognitive assessment scores and traditional visual image analysis. This could be particularly useful for monitoring the progression of patients with mild cognitive impairment (MCI), which is often a transitional stage between the cognitive decline associated with normal ageing, and that of established AD.

The neuropathological changes associated with the development of AD begin many years before cognitive symptoms become apparent. According to the amyloid cascade hypothesis, the disease process begins with the formation of insoluble β-amyloid (Aβ) plaques, whose presence triggers the hyperphosphorylation of tau protein, ultimately leading to cell death (Selkoe, 1991). Surrogate measures of the levels of Aβ and tau in the brain may be obtained from the cerebrospinal

* Corresponding author.
E-mail address: krg03@imperial.ac.uk (K.R. Gray).

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.09.065

fluid (CSF). Various studies have shown reduced CSF Aβ and elevated CSF tau in AD patients compared with cognitively normal individuals (for example, Motter et al., 1995; Vandermeeren et al., 1993). MCI patients tend to have CSF Aβ and tau levels between those expected of AD patients and healthy controls, with AD-like biomarker levels associated with an increased likelihood of progression to AD (Hansson et al., 2006). Positron emission tomography (PET) imaging with radiotracers for amyloid provides an alternative method for assessing intracranial Aβ deposition (Klunk et al., 2004).

PET imaging with the radiotracer [^{18}F]-fluorodeoxyglucose (FDG) can be used to assess brain function in terms of the rate of cerebral glucose metabolism. Both AD and MCI are associated with significantly reduced glucose metabolism in affected regions, including temporal and parietal lobes, and the posterior cingulate cortex (Herholz et al., 2002; Langbaum et al., 2009). Reduced metabolism in AD patients is predictive of their cognitive decline and histopathological diagnosis (Hoffman et al., 2000; Silverman et al., 2001), and in MCI patients can predict their progression to AD (Anchisi et al., 2005). Changes in metabolism can be detected on FDG-PET before corresponding structural changes are visible (Aisen et al., 2010). FDG-PET-based classification techniques can discriminate both AD and MCI patients from healthy controls (Hinrichs et al., 2009).

The progressive structural damage caused by AD can be non-invasively assessed by using magnetic resonance imaging (MRI) to measure cerebral atrophy or ventricular expansion. Temporal lobe atrophy is closely associated with AD, and histological studies show that the hippocampus, amygdala and entorhinal cortex are particularly vulnerable to AD pathology (Braak and Braak, 1998). Accelerated hippocampal atrophy compared with healthy controls has been measured using MRI in both AD and MCI patients (Schuff et al., 2009; van de Pol et al., 2007). A recent study (Cuingnet et al., 2011) comparing ten MRI-based high-dimensional classification methods reported good results in discriminating between AD patients and healthy controls (up to 81% sensitivity and 95% specificity). Two methods relied on only hippocampal shape or volume, while the remainder were whole-brain approaches, using either cortical thickness measures, or voxel-wise tissue class probabilities. Four of the ten methods achieved slightly better accuracy than a random classifier in distinguishing MCI patients who later progressed to AD from MCI patients remaining stable over 18 months.

Although age is the most significant risk factor for AD (Rocca et al., 1991), genetic and environmental factors also play a role. The ApoE gene is the only one so far shown to be associated with sporadic late-onset AD (Dawbarn and Allen, 2007). There are three major alleles of this gene: ε2, ε3 and ε4. The most common is ε3, while ε4 is associated with an increased risk of developing AD, and ε2 with a reduced risk (Corder et al., 1993). More extensive AD pathology is generally observed in carriers of the ApoE ε4 allele than in non-carriers (Roses and Saunders, 1997). Genetics can therefore impact the biological and neuroimaging biomarkers. For example, AD carriers of the ApoE ε4 allele typically have reduced CSF Aβ, elevated CSF tau, and accelerated hippocampal atrophy on MRI compared with non-carriers (Schuff et al., 2009; Tapiola et al., 2000). Cognitively normal carriers of the ApoE ε4 allele display reduced glucose metabolism on FDG-PET in AD-typical regions (Langbaum et al., 2009).

The above biomarker patterns, however, are not specific to AD. For example, reduced CSF Aβ and elevated CSF tau are also associated with Lewy body dementia and vascular dementia (Andreasen et al., 2001; Blennow and Hampel, 2003). Temporal lobe atrophy is also observed in hippocampal sclerosis and temporal lobe epilepsy (Keihaninejad et al., 2012), and an AD-like metabolic pattern can also indicate Creutzfeldt–Jakob disease (Hoffman et al., 1990). It has been suggested that there may be some complementary information between modalities which can be exploited to produce more powerful combined classifiers (Landau et al., 2010; Walhovd et al., 2010). The potential utility of biomarker combinations is also supported by the recently revised diagnostic criteria (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011). For research applications, these incorporate a biomarker-based probability of AD aetiology which is highest when evidence of both amyloid deposition and neuronal degeneration or injury are observed. This support has resulted in an increasing interest in using multi-modality imaging and biological data for classification. Studies reporting multi-modality AD classification are far less common than those based on data from a single modality (Weiner et al., 2012). However, two independent studies (Hinrichs et al., 2011; Zhang et al., 2011) using multi-kernel learning report that classification based on multi-modality data is superior to that based on any individual modality. Classification accuracies of 93% between AD patients and healthy controls, and 76% between MCI patients and healthy controls were reported in Zhang et al. (2011), based on a combination of FDG-PET, MRI and CSF measures.

We present a framework for multi-modality classification based on pairwise similarity measures derived from random forests (Breiman, 2001). The similarities are used to construct a manifold representation from labelled training data and then to infer the clinical labels of test data mapped into this space. Classical multidimensional scaling (MDS) (Torgerson, 1952) is applied to learn a manifold on which to perform classification, resulting in the construction of a coordinate embedding in which the distances between points preserve their pairwise similarities. MDS is commonly used to provide low-dimensional visualisations of similarity relationships, including those derived from random forests (Hastie et al., 2011). Random forest-derived similarities have been successfully applied in unsupervised clustering tasks, for example those involving high-dimensional genetic or tissue microarray data (Shi and Horvath, 2006; Shi et al., 2005). Here, random forests are used to derive *supervised* similarity measures, with the aim of generating manifolds that are optimal for the task of clinical group discrimination. The proposed method facilitates the incorporation of multi-modality data, since similarities derived from several datasets may be combined to generate an embedding that simultaneously encodes information from all features.

Manifold learning techniques based on pairwise similarities between images have been applied in a variety of neuroimaging studies. For example, Laplacian eigenmaps (Belkin and Niyogi, 2003) have been used to generate an embedding of brain MR images based on similarities derived from overlaps of their structural segmentations (Aljabar et al., 2008). Isomap (Tenenbaum et al., 2000) has also been used to estimate the manifold structure of brain MR images, using distance measures based on nonrigid transformations between image pairs (Gerber et al., 2010). A framework for fusing manifold learning steps based on different pairwise similarity measures was presented in Aljabar et al. (2010). The method proposed here uses random forests to derive consistent pairwise similarity measures for multiple modalities, thus facilitating the combination of different types of feature data. We evaluate the proposed multi-modality classification framework by application to neuroimaging and biological data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). These data include FDG-PET and MR imaging data, CSF biomarker measures, and categorical genetic information. This work is an extension of preliminary studies involving neuroimaging data which have been previously presented at a workshop (Gray et al., 2011).

Methodology

A schematic overview of the proposed approach is illustrated in Fig. 1. A random forest classifier was applied to the feature data from each modality independently, not only to obtain single-modality classification results for comparison, but also to derive the similarities required for manifold learning. The resulting similarity matrices were combined, and classical MDS was applied to generate a joint embedding for multi-modality classification. Full details of the data collection and feature extraction are presented in the Neuroimaging and biological feature data section.

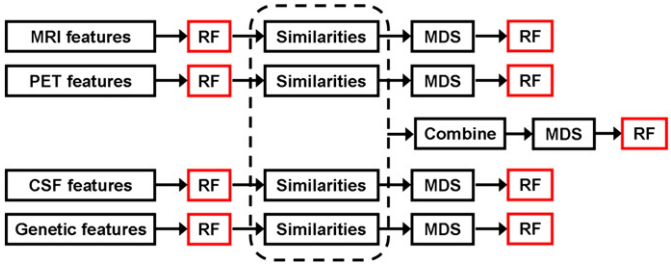

Fig. 1. Schematic overview of the proposed methodology. Each random forest (RF) step provides a classification result whose performance will be reported. Random forests are used both to derive the pairwise similarity measures for each feature set, and also to perform the single- and multi-modality classification experiments.

Random forests for classification

A random forest (Breiman, 2001) is an ensemble classifier consisting of many decision trees, where the final predicted class for a test example is obtained by combining the predictions of all individual trees, as illustrated in Fig. 2. Random forests combine bootstrap aggregation (bagging) (Breiman, 1996) and random feature selection (Amit and Geman, 1997; Ho, 1998) to construct a collection of decision trees exhibiting controlled variation. We used the R implementation of random forests, a port of Leo Breiman and Adele Cutler's original Fortran code by Andy Liaw and Matthew Wiener, version 4.6–6 (http://cran.r-project.org/web/packages/randomForest).

The training set for each individual tree in a random forest is constructed by sampling N examples at random with replacement from the N available examples in the dataset. This is known as bootstrap sampling, and bagging describes the aggregation of predictions from the resulting collection of trees. As a result of the bootstrap sampling procedure, approximately one third of the available N examples are not present in the training set of each tree. These are referred to as the "out-of-bag" data of the tree, for which internal test predictions can be made. By aggregating the predictions of the out-of-bag data across all trees, an internal estimate of the generalisation error of the random forest can be determined.

At each node in a tree, $d \ll D$ features are randomly selected from the D available features in the dataset, and the node is partitioned using the best possible binary split. A parent node n_p is partitioned into child nodes n_l and n_r according to the Gini index (Breiman et al., 1984), which measures the likelihood that an example would be incorrectly labelled if it were randomly classified according to the distribution of labels within the node. For a binary split, the Gini index of a node n may be expressed as $I_G(n) = 1 - \sum_{c=1}^{2} p_c{}^2$, where p_c is the relative proportion of examples belonging to class c present in node n. The best possible binary split is the one which maximises the improvement in the Gini index $\Delta I_G(n_p) = I_G(n_p) - p_l I_G(n_l) - p_r I_G(n_r)$, where p_l and p_r are the proportions of examples in node n_p that are assigned to child nodes n_l and n_r, respectively. The Gini index can also be used to assess the relative importance of features for classification. A measure of the importance of an individual feature may be computed by summing the decreases in the Gini index occurring at all nodes in the forest which are partitioned based on that feature.

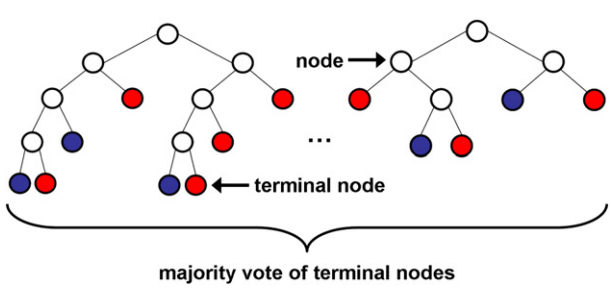

Fig. 2. Illustration of a random forest, showing two trees in detail. Each node is partitioned based on a single feature, and each branch ends in a terminal node. Terminal nodes provide a prediction for the class of a test example based on the path taken through the tree. The colour of a terminal node indicates its class prediction. The final predicted class for a test example is obtained by combining the predictions of all individual trees.

Manifold learning based on random forest similarities

Random forests can provide measures of the similarity between pairs of examples in the dataset. Each of the N examples is represented by a feature vector, all of which are passed down each tree in the forest. The similarities are initialised to zero, and if examples i and j finish in the same terminal node of a tree, their similarity s_{ij} is increased by one. The final pairwise similarity measures are normalised by the total number of trees in the forest.

The similarities thus form a $N \times N$ matrix with elements s_{ij}, and corresponding distance matrix elements $d_{ij} = 1 - s_{ij}$ (Cox and Cox, 2001). Manifold learning techniques may be applied to find an appropriate coordinate embedding for the feature vectors, such that the distance relationships between them are preserved. A review of the most popular manifold learning techniques, as applied to medical imaging, is provided in Aljabar et al. (2012). MDS (Torgerson, 1952) is commonly used in conjunction with random-forest derived similarity measures for low-dimensional data visualisation, as well as unsupervised clustering tasks. In this work, MDS is applied to similarity measures derived in a supervised manner from random forests, with the aim of generating manifolds that are optimal for the task of clinical group discrimination.

Using MDS, the matrix of coordinates $\mathbf{X}$ is derived by performing an eigenvalue decomposition on the matrix of scalar products

$$\mathbf{B} = \mathbf{X}^{\mathbf{T}}\mathbf{X} = -\frac{1}{2}\left\{ d_{ij}^2 - \frac{1}{N}\sum_{i=1}^{N} d_{ij}^2 - \frac{1}{N}\sum_{j=1}^{N} d_{ij}^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}^2 \right\}.$$

Retaining only the eigenvectors corresponding to the k largest-valued eigenvalues leads to a k-dimensional embedding for the data. A goodness-of-fit parameter G, describing the extent to which the selected k eigenvectors represent the full matrix, can be useful in selecting an appropriate dimensionality for the embedding. The measure of goodness-of-fit used in this work is given by

$$G = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{N} \max\left(\lambda_j, 0\right)},$$

where the eigenvalues λ_j are sorted in decreasing order (Mardia et al., 1979).

To generate an embedding that simultaneously incorporated information from multiple modalities, a joint similarity matrix $\mathbf{S}$ was defined as a linear combination of the similarity matrices from each of the four modalities $\mathbf{S_i}$. Each modality was assigned a weighting factor α_i, such that $\mathbf{S} = \sum_{i=1}^{4} \alpha_i \mathbf{S_i}$, where $\sum_{i=1}^{4} \alpha_i = 1$. To ensure the best combination of the four modalities for classification, the α_i parameters were optimised as part of the training process. This was achieved by performing a grid-search within the training data to select the optimum modality weightings. The classifier was then trained using the selected set of parameters, before having its performance assessed on the test data.

Experiments and results

We applied the proposed methodology to neuroimaging and biological data from 147 ADNI participants for whom baseline 1.5 T MRI, FDG-PET and CSF biomarker measures were available, as well as ApoE genotype information.

Neuroimaging and biological feature data

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.ucla.edu). The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. A more detailed description of the ADNI study is provided in Appendix A.

We use imaging and biological data from 147 ADNI participants, including 37 AD patients, 75 MCI patients, and 35 healthy controls (HC). These represent all participants for whom ApoE genotype information and baseline 1.5 T MRI, FDG-PET and CSF measures of Aβ, tau and phosphorylated tau (ptau) were available, subject to the exclusions described in Appendix B. The MCI patients were divided into those who have progressed to AD (pMCI) and those whose diagnoses have so far remained stable (sMCI) based on changes in clinical status occurring over 20 ± 11 months (range 6–36). Clinical and demographic information for the study population is provided in Table 1.

MRI feature extraction

Automatic whole-brain segmentations into 83 anatomical regions were prepared in native MRI space using multi-atlas propagation with enhanced registration (MAPER), an approach that has been previously described and validated for use in AD (Heckemann et al., 2010). The segmentations are available to download through the ADNI website, and full details of the procedure and morphometric analysis are presented in Heckemann et al. (2011). The required atlas data consisted of manually segmented T1-weighted MR volumes from 30 young, healthy adults, as described in Hammers et al. (2003). Protocols for the manual delineation are described in Hammers et al. (2003) and Gousias et al. (2008).

To match the requirements of MAPER, additional pre-processing was applied for brain extraction and tissue classification. For brain extraction, intracranial masks were generated using binary masks covering intracranial white and grey matter as the starting point, as described in Heckemann et al. (2011). The binary masks had been generated with a semi-automatic procedure as part of a separate project using MIDAS (Freeborough et al., 1997; Leung et al., 2011). Individual tissue probability maps for grey matter, white matter and CSF were obtained using FSL FAST (http://www.fmrib.ox.ac.uk/fsl). Binary maximum-probability tissue class labels were employed to mask the anatomical segmentations. All regions except ventricles, central structures, cerebellum and brainstem were masked with a grey matter label. Lateral ventricles were masked with a CSF label. Regional volumes were normalised by the intracranial volume, resulting in 83 volumetric region-based features per image.

FDG-PET feature extraction

Each FDG-PET image was motion-corrected as necessary, converted to a 30-minute static image, examined for major artefacts, and affinely aligned with the corresponding MRI using the Image Registration Toolkit (IRTK; http://www.doc.ic.ac.uk/dr/software). An affine transformation was preferred over a rigid one because it can account for scaling or voxel size errors remaining after phantom correction of the MRI (Clarkson et al., 2009).

Using the "Segment" module of SPM5 (http://www.fil.ion.ucl.ac.uk/spm), each MRI was linearly and non-linearly deformed to the Montreal Neurological Institute (MNI) template (Ashburner and Friston, 2005). The resulting transformation parameters were applied to the MRI-space PET images using trilinear interpolation. The MNI-space PET images (voxel sizes $2 \times 2 \times 2$ mm) were then smoothed to a common isotropic spatial resolution of 8 mm full-width-at-half-maximum (FWHM) using scanner-specific kernels (Joshi et al., 2009), and then by an additional 8 mm FWHM isotropic Gaussian kernel. The smoothed images were intensity normalised to account for inter-subject variability in overall radioactivity using an independently-derived cluster of relatively preserved regions (Yakushev et al., 2009). The *a priori* brain mask available in SPM5 was thresholded at 50% probability, and applied to each normalised FDG-PET image to exclude voxels outside the brain. Signal intensities were extracted from all remaining voxels, resulting in 239,304 voxel-based features per image.

Biological features

The ADNI Biomarker Core provides biological data for the study participants. These data include CSF measures of Aβ, tau and ptau, as well as ApoE genotype information determined from a blood sample. Details of the biofluid collection and processing are provided in Trojanowski et al. (2010). The genetic feature data for each participant consisted of a single categorical variable describing their ApoE genotype. These data are summarised in Table 2.

Classification experiments

Classification performance was assessed between three clinically relevant pairs of diagnostic groups (AD/HC, MCI/HC, pMCI/sMCI). Robust estimates of classifier performance were obtained using a stratified repeated random sampling approach. The mean accuracy, balanced accuracy, sensitivity and specificity were evaluated over 100 runs in which 75% of the data were randomly selected for training, with the remaining 25% used as test data. Balanced accuracy is the arithmetic mean of the sensitivity and specificity. This provides a more meaningful performance metric for groups of unequal sizes.

Single-modality classification results

A random forest classifier was applied to the feature data from each of the four modalities independently, and the single-modality classification results obtained are presented in Table 3.

Before performing classification experiments, the number of trees grown in each forest, t, and the number of features randomly selected at each tree node, d, had to be chosen. We used $t = 5000$ for all the experiments that follow, since stable estimates of the out-of-bag classification error were consistently observed for $t \gtrsim 1000$. We used $d = \sqrt{D}$ for all experiments, following the recommendation of Liaw and Wiener (2002). The value of d was also consistently observed to have little effect on the out-of-bag classification error estimate.

As described in the Random forests for classification section, estimates of the relative importance of the features for classification may

Table 1
Clinical and demographic information for the study population. For each group, the total number of subjects (N) and number of females (F) are shown, along with the average age, average mini mental state examination (MMSE) score, and clinical dementia rating (CDR).

	N(F)	Age	MMSE score	CDR (%)		
		(mean ± std. dev.)	(mean ± std. dev.)	0	0.5	1
AD	37(14)	76.8 ± 6.6	23.5 ± 2.0	0	49	51
pMCI	34(12)	76.1 ± 7.3	26.5 ± 1.7	0	100	0
sMCI	41(12)	75.7 ± 6.5	27.3 ± 1.8	0	100	0
HC	35(12)	74.5 ± 5.2	28.9 ± 1.2	100	0	0

Table 2
Summarised biological and genetic information for the study population. CSF measures are expressed as mean ± std. dev., and genetic information is expressed as the proportion of each diagnostic group possessing each of the five ApoE allele pairs.

	CSF measures (pg/mL)			ApoE genotype (%)				
	Aβ	tau	ptau	ϵ_3, ϵ_3	ϵ_3, ϵ_4	ϵ_4, ϵ_4	ϵ_2, ϵ_3	ϵ_2, ϵ_4
AD	141 ± 43	127 ± 64	46 ± 24	24	41	30	0	5
pMCI	152 ± 47	96 ± 41	37 ± 12	41	41	18	0	0
sMCI	167 ± 55	105 ± 81	35 ± 22	39	39	12	10	0
HC	208 ± 56	69 ± 28	26 ± 16	48	23	0	26	3

Table 3
Single-modality classification accuracy (acc.), balanced accuracy (bacc.), sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the original imaging and biological feature data. Results are expressed as mean (standard error).

		CSF	MRI	FDG-PET	Genetic
AD/HC	Acc. (%)	76.5 (0.8)	81.6 (0.8)	86.0 (0.7)	72.6 (0.9)
	Bacc. (%)	76.8 (1.3)	81.8 (1.3)	86.0 (1.2)	72.7 (1.3)
	Sens. (%)	73.0 (1.3)	79.8 (1.3)	86.8 (1.1)	71.3 (1.3)
	Spec. (%)	80.5 (1.3)	83.8 (1.3)	85.1 (1.3)	74.1 (1.4)
MCI/HC	Acc. (%)	63.1 (0.8)	66.9 (0.9)	66.5 (0.8)	73.8 (0.5)
	Bacc. (%)	63.8 (1.4)	68.9 (1.3)	66.9 (1.3)	60.7 (0.9)
	Sens. (%)	62.0 (1.1)	63.7 (1.2)	65.7 (1.1)	94.7 (0.5)
	Spec. (%)	65.5 (1.7)	74.0 (1.4)	68.1 (1.5)	26.6 (1.2)
pMCI/sMCI	Acc. (%)	52.9 (1.0)	55.1 (1.0)	52.6 (1.0)	47.3 (0.9)
	Bacc. (%)	53.5 (1.6)	55.5 (1.7)	53.1 (1.7)	42.4 (2.4)
	Sens. (%)	58.1 (1.6)	59.1 (1.8)	57.4 (2.0)	32.1 (2.1)
	Spec. (%)	48.8 (1.5)	51.9 (1.6)	48.8 (1.5)	52.6 (2.7)

be extracted from the random forest. Feature importance for the two imaging modalities are shown in Fig. 3. The most important features for MRI include volumes of the hippocampus, amygdala, and other medial temporal lobe structures. The most important features for FDG-PET include signal intensities of voxels located in the posterior cingulate gyrus, parietal lobe, posterior temporal lobe, and around the hippocampus.

Single-modality similarity-based classification results

The random forest classifiers described in the Single-modality classification results section were used to derive pairwise similarity measures for each of the four modalities, as described in the Manifold learning based on random forest similarities section. Examples of the similarity matrices are shown in Fig. 4.

MDS was applied to each similarity matrix, and a goodness-of-fit value of 90% was used to determine an appropriate dimensionality for the resulting embeddings. A random forest classifier was then applied to the embedded feature data from each of the four modalities independently. The single-modality classification results obtained are presented in Table 4, along with the dimensionality of each embedding.

No consistent differences are observed between the balanced accuracies based on the embedded feature data shown in Table 4, and those based on the original feature data shown in Table 3.

Multi-modality similarity-based classification results

MDS was applied to the joint similarity matrix constructed using information from all four modalities, and a goodness-of-fit value of 90% was again used to determine an appropriate dimensionality for the resulting embedding. A random forest classifier was applied to the embedded feature data, and the multi-modality classification results obtained are presented in Table 5. The balanced accuracies based on multi-modality classification are significantly ($p<0.01$) higher than those based on any individual modality for both the AD/HC and MCI/HC experiments. For the pMCI/sMCI experiment, however, the balanced accuracy based on multi-modality classification is not significantly different from that based on MRI information alone. Table 5 additionally shows results based on the application of a random forest classifier to the combined feature set generated by a simple concatenation of features from all four modalities. These show no significant differences from the results based on FDG-PET alone.

For each of the classification experiments, the distribution of modality weighting parameters selected over the 100 runs is illustrated in Fig. 5.

Discussion and conclusion

We have presented a framework for multi-modality classification based on pairwise similarity measures derived from random forests. The similarities are used to construct a manifold representation from labelled training data and then to infer the clinical labels of test data mapped into this space. Random forests are used to derive similarity measures in a supervised manner, with the aim of generating manifolds that are optimal for the task of clinical group discrimination. The proposed method facilitates the incorporation of multi-modality data, since consistent similarities may be derived from several datasets, and combined to generate an embedding that simultaneously encodes information from all features. Multi-modality classification may then be performed using coordinates from this joint embedding. The applicability of the method to a large and diverse dataset is demonstrated using imaging and biological information from the ADNI study. This includes FDG-PET and MR imaging data, CSF biomarker measures, and categorical genetic information.

Classification based on the joint embedding constructed using information from all four modalities is superior to the classification based on any individual modality for comparisons between AD patients and HC, as well as between MCI patients and HC. This finding is in agreement with other recent studies (Hinrichs et al., 2011; Zhang et al., 2011), and supports previous suggestions that there is some complementary information between modalities which can be exploited to produce more powerful combined classifiers (Landau et al., 2010; Walhovd et al., 2010). Classification performance is commonly reported in terms of accuracy, but here we perform statistical comparisons between experiments based on the balanced accuracy. This provides a more meaningful performance metric for groups of unequal sizes. In terms of accuracy, we achieve 89% classification between AD patients and HC, and 75% between MCI patients and HC.

Direct comparisons with existing work are complicated by factors such as the inclusion of different subjects and modalities, as well as the use of different methods for feature extraction and cross-validation. In assessing the proposed framework, we consider comparisons between the single- and multi-modality classification performance more important, since these are based on the same method, data, and processing pipeline. Our multi-modality classification results

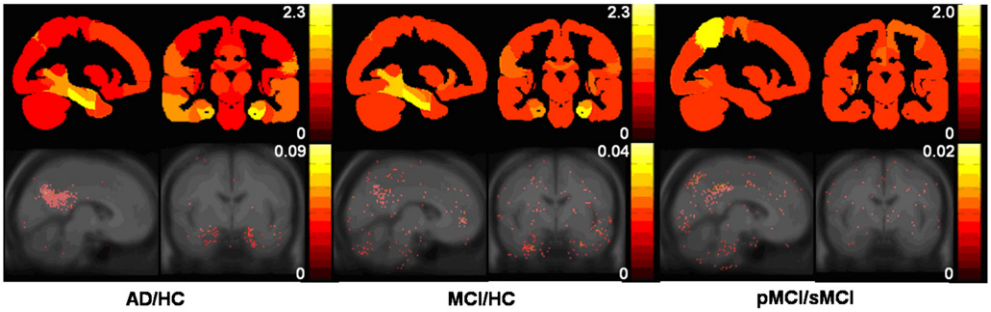

Fig. 3. Feature importance for discriminating between the three clinical group pairs using region-based MRI (top), and voxel-based FDG-PET (bottom). For MRI, regional feature importance is superimposed onto slices of a maximum probability brain atlas which has been masked in the same way as the anatomical segmentations. For FDG-PET, important voxels are overlaid onto a MNI-space average MR image.

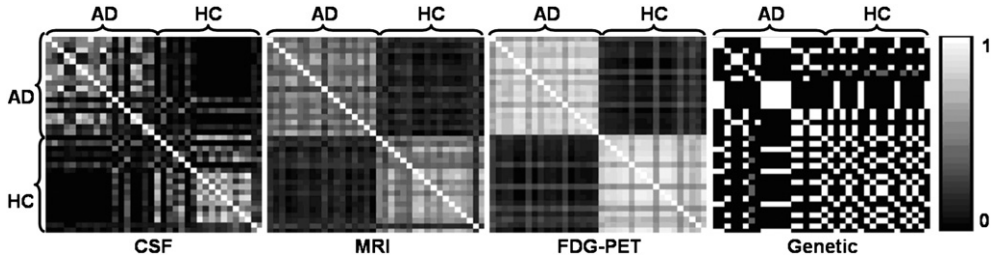

Fig. 4. Similarity matrices for each of the four modalities for the AD/HC experiment. Matrices are symmetric, and each entry represents the similarity between a pair of subjects based on the input feature data.

are, however, comparable with those reported in other recent studies based on ADNI data. Zhang et al. (2011) combine MRI, FDG-PET and CSF data to achieve classification accuracies of 93% between AD patients and HC, and 76% between MCI patients and HC. Hinrichs et al. (2011) combine MRI, FDG-PET, CSF, genetic and cognitive data to achieve an accuracy of 92% between AD patients and HC. Both studies employ multi-kernel classifiers based on support vector machines, which are relatively well-established in the field (Weiner et al., 2012). Random forests have not been extensively applied in neuroimaging research, although they are becoming more widely used for applications including classification (Chincarini et al., 2011; Hope et al., 2008) and segmentation (Geremia et al., 2011; Iglesias et al., 2011). Classification of neurodegenerative disease based on the combination of imaging and non-imaging information is therefore a new application for random forests, and we are encouraged by the fact that our results are comparable with those of more established techniques.

The ability of random forests to extract consistent pairwise similarity measures for multiple modalities facilitates the combination of different types of feature data. We demonstrate this using datasets in which the number of features differs between modalities by several orders of magnitude. A simple concatenation of all features was shown not to optimally combine these data (Table 5), since the high-dimensional FDG-PET features dominate the results.

In a previous work (Gray et al., 2011), we obtained comparable multi-modality classification results using only information extracted from the two neuroimaging modalities (accuracies of 90% for AD/HC, and 76% for MCI/HC). The lack of improvement over these previously reported results is likely to be attributable to the considerable reduction in size of the subject group as a result of our requirement for CSF biomarker information. The subject group for this study is approximately half the size of that used in Gray et al. (2011). The present work additionally employs a more robust form of cross-validation, using a stratified repeated random sampling approach, as opposed to the single round of ten-fold cross-validation employed in Gray et al. (2011).

In the context of a neuroimaging application, estimates of the importance of the features for classification are valuable because this allows assessment of whether the features that contribute most to the classifier correspond to regions or structures with a biologically plausible connection to pathology. In this work, the most important features for discriminating between clinical groups correspond with those known to be visibly affected in AD on both FDG-PET and structural MR imaging (Hampel et al., 2008; Patwardhan et al., 2004). Important features for distinguishing between AD patients and HC are more localised to affected areas, with the more challenging distinctions between MCI patients and HC, or pMCI and sMCI patients, requiring features spread across a larger part of the brain. The motivation for extracting region-based features from the MR images and voxel-based features from the FDG-PET images was to demonstrate that these two different types of imaging features could be readily combined using the proposed method.

The lack of significant difference between classification performance based on the original feature data and that based on the embedding coordinates for each individual modality is expected, since a random forest is already a nonlinear classifier. However, a difference is observed for the comparison between MCI patients and HC based on the voxel-wise FDG-PET features. This may be due to the inhomogeneity of the MCI group, which comprises both pMCI and sMCI patients. It is possible that the high-dimensional voxel-based FDG-PET features are sensitive to differences in the pattern of glucose metabolism between these two groups, resulting in a reduced classification performance based on the associated embedding coordinates.

Random forests are ensemble-based classifiers that are often applied to high-dimensional datasets. In this work, random forests are also applied to low-dimensional biological data so that consistent pairwise

Table 4
Single-modality classification accuracy (acc.), balanced accuracy (bacc.) sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the embedded imaging and biological feature data. Results are expressed as mean (standard error). The mean dimensionality of each embedding (k) is also shown.

		CSF	MRI	FDG-PET	Genetic
AD/HC	Acc. (%)	76.1 (0.8)	82.5 (0.7)	86.4 (0.7)	72.6 (0.9)
	Bacc. (%)	76.3 (1.3)	82.1 (1.4)	86.5 (1.2)	72.7 (1.3)
	Sens. (%)	72.8 (1.3)	88.6 (1.2)	85.8 (1.2)	71.3 (1.3)
	Spec. (%)	79.8 (1.4)	75.6 (1.5)	87.1 (1.3)	74.1 (1.4)
	k	13	22	9	2
MCI/HC	Acc. (%)	61.7 (0.8)	67.3 (1.0)	53.5 (0.7)	73.8 (0.5)
	Bacc. (%)	61.7 (1.3)	69.1 (1.4)	60.2 (1.2)	60.7 (0.9)
	Sens. (%)	61.6 (1.1)	64.3 (1.3)	42.3 (1.1)	94.7 (0.5)
	Spec. (%)	61.8 (1.5)	73.9 (1.4)	78.0 (1.3)	26.6 (1.2)
	k	25	47	35	2
pMCI/sMCI	Acc. (%)	52.1 (1.0)	58.4 (1.0)	53.0 (1.0)	43.5 (0.9)
	Bacc. (%)	52.7 (1.7)	58.3 (1.7)	52.8 (1.7)	41.2 (2.4)
	Sens. (%)	57.9 (1.6)	56.9 (1.6)	50.6 (1.8)	27.4 (2.0)
	Spec. (%)	47.5 (1.7)	59.7 (1.8)	54.9 (1.6)	55.0 (2.7)
	k	21	38	35	1

Table 5
Multi-modality classification accuracy (acc.), balanced accuracy (bacc.) sensitivity (sens.) and specificity (spec.) based on the application of a random forest classifier to the jointly embedded imaging and biological feature data, as well as the concatenated set of all features. Results are expressed as mean (standard error). The mean dimensionality of each embedding (k) is also shown.

		Combined embedding	Concatenated features
AD/HC	Acc. (%)	89.0 (0.7)	86.2 (0.7)
	Bacc. (%)	89.0 (1.2)	87.1 (1.1)
	Sens. (%)	87.9 (1.2)	85.1 (1.4)
	Spec. (%)	90.0 (1.1)	86.1 (1.3)
	k	18	–
MCI/HC	Acc. (%)	74.6 (0.8)	66.3 (0.8)
	Bacc. (%)	72.7 (0.8)	65.3 (1.1)
	Sens. (%)	77.5 (1.0)	68.5 (1.5)
	Spec. (%)	67.9 (1.7)	66.9 (1.3)
	k	20	–
pMCI/sMCI	Acc. (%)	58.0 (0.9)	53.0 (1.1)
	Bacc. (%)	57.9 (1.7)	57.3 (1.9)
	Sens. (%)	57.1 (1.8)	49.6 (1.4)
	Spec. (%)	58.7 (1.5)	53.5 (1.7)
	k	29	–

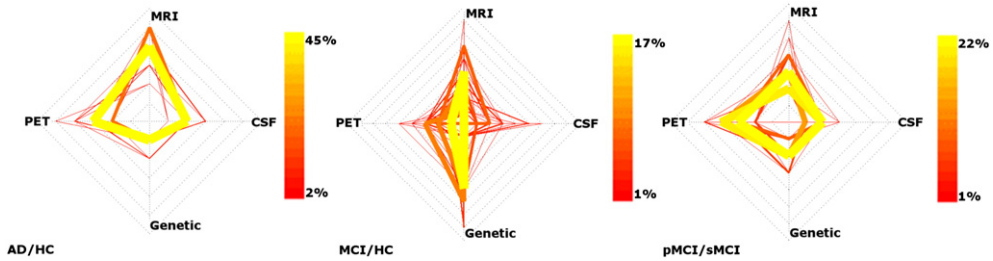

Fig. 5. Cobweb plots showing the distribution of parameters selected over the 100 leave-25%-out runs for all three classification experiments. The four spokes of each plot represent the four modalities, and each coloured line connecting the four spokes represents a set of parameter values. The colour and weight of each line represents the percentage of runs in which the associated parameter set was selected.

similarity measures may be obtained for all modalities. In the case of a single feature, such as the categorical genetic information, a random forest reduces to bootstrap aggregation.

Visualisation of the parameters selected to combine similarities for multi-modality classification (Fig. 5) provides some interesting insights into the relationships among the modalities. The figure indicates the optimum way in which to combine MRI, FDG-PET, CSF and genetic information within the framework described. For distinguishing between AD patients and HC, for example, it appears that FDG-PET and MR imaging features provide the most complementary information. For distinguishing MCI patients from HC, genetic information appears to have a relatively high importance. However, these indications may not necessarily have a direct biological interpretation, and are likely not sufficient to draw any firm conclusions about the relative importance of the four modalities. The optimum modality weightings for distinguishing between AD patients and HC are more stable than for the other two experiments. The weighting parameters selected for distinguishing MCI patients from HC maybe more variable because the heterogeneity of the MCI group makes their selection dependent on the proportions of pMCI and sMCI patients in the training set. The selected modality weightings are not equal for any experiment, supporting the need to optimise the relative contributions of the four modalities for classification. The figure suggests an interesting avenue for further research, in that estimates of inter-modality correlations could help to determine the amount of complementary information between them. This could facilitate decisions on how to acquire the maximum amount of diagnostically relevant information for a patient using a minimum number of assessments.

The classification performance between pMCI and sMCI patients is not significantly improved by combining multi-modality information in this study. Other studies which apply multi-modality classification to this challenging problem have incorporated longitudinal information to improve performance (Hinrichs et al., 2011; Zhang et al., 2012). Our previous work based on FDG-PET (Gray et al., 2012) has also shown that incorporating longitudinal information can be beneficial to improve the ability to distinguish between these two groups. This will be an important issue to address in the future. It is important to consider, however, that progression from MCI to AD occurs at a rate of 10–15% per year (Petersen et al., 1999), with up to 80% of MCI patients developing AD over a six year period (Petersen, 2004). Longer clinical follow-up is therefore required to properly assess the utility of any classification method in separating pMCI from sMCI patients. This may be made possible by the continuation of the ADNI study in the form of ADNI-GO and ADNI-2. Further information about these studies is available via the ADNI website (http://adni.loni.ucla.edu/about/about-the-study/).

We have identified several areas for further research. Methodologically, the approach is generalisable, in that the manifold learning step could be performed using an alternative technique, and similarities could be combined using more sophisticated formulae. We applied MDS for manifold learning in this work because its application to random forest-derived similarities is straightforward, and relatively common in the literature (Hastie et al., 2011). However, MDS may not be the optimal embedding approach, and there is potential to improve the performance of this step by pursuing a thorough exploration of alternative techniques. Similarly, we chose to additively combine similarity matrices for simplicity, and there is potential for improvement by incorporating a more sophisticated formula. Random forest classifiers extend naturally to multi-class problems, and the framework described here could be applied to distinguish between multiple patient groups in the future. Exploration of the multi-class setting would better represent a real-world clinical scenario, including patients with other forms of dementia. In addition, the implementation of random forests used in this work could be modified to produce uncertainty information about the predicted diagnostic labels. This may be more useful to clinicians than a simple binary prediction, and could be achieved by having each leaf node store a probabilistic distribution of labels, rather than a point estimate. Criminisi et al. (2012), for example, describes this and other extensions to the original random forests algorithm, and presents a unified model of random decision forests for classification, regression, density estimation, manifold learning, and semi-supervised learning. Further investigations into different data fusion strategies may also be interesting. Although the feature level fusion approach proposed in this work can capture interdependencies between modalities, the application of a data level fusion technique to the neuroimaging data may be advantageous in terms of capturing local inter-modality correlations.

Acknowledgments

The authors would like to thank Igor Yakushev (University of Mainz, Germany) for provision of the reference cluster image used for FDG-PET intensity normalisation, as well as Casper Nielsen and Kelvin Leung (Dementia Research Centre, University College London) for provision of the MIDAS brain masks used for MRI anatomical segmentation. Katherine R. Gray received a studentship from the Engineering and Physical Sciences Research Council. Rolf A. Heckemann was supported by a research grant from the Dunhill Medical Trust. The Image Registration Toolkit was used under licence from Ixico Ltd.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The

grantee organisation is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Appendix A. The Alzheimer's Disease Neuroimaging Initiative

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organisations, as a $60 million, five-year public–private partnership. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. Subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit approximately 200 cognitively normal older individuals to be followed for three years, 400 MCI patients to be followed for three years, and 200 early AD patients to be followed for two years. Further up-to-date information is available on the ADNI information website (http://www.adni-info.org).

Appendix B. Exclusions

In this work, we use imaging and biological data from 147 ADNI participants. These represent all participants for whom ApoE genotype information and baseline 1.5 T MRI, FDG-PET and CSF measures of Aβ, tau and phosphorylated tau were available, subject to the following exclusions. FDG-PET images acquired using either the Siemens HRRT or BioGraph HiRez scanners were excluded due to differences in the pattern of FDG metabolism that were discovered during the ADNI quality control process. Further information is available on the ADNI PET Core website (http://www.loni.ucla.edu/twiki/bin/view/ADNI/ADNIPETCore). A small number of subjects ($n = 14$) whose imaging data could not be processed as required, or whose diagnosis did not clearly fall into one of the four diagnostic groups (AD, pMCI, sMCI, HC), were also excluded. ADNI subject identifiers for these participants are provided as supplementary material, along with more detailed reasons for their exclusion.

Appendix C. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.09.065.

References

Aisen, P.S., Petersen, R.C., Donohue, M.C., Gamst, A., Raman, R., Thomas, R.G., Walter, S., Trojanowski, J.Q., Shaw, L.M., Beckett, L.A., Jack Jr., C.R., Jagust, W., Toga, A.W., Saykin, A.J., Morris, J.C., Green, R.C., Weiner, M.W., the Alzheimer's Disease Neuroimaging Initiative, 2010. Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans. Alzheimers Dement. 6, 239–246.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 270–279.

Aljabar, P., Rueckert, D., Crum, W.R., 2008. Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. Neuroimage 43, 225–235.

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Boardman, J.P., Murgasova, M., Doria, V., Rutherford, M., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2010. Combining morphological information in a manifold learning framework: application to neonatal MRI. 13th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'10), pp. 1–8.

Aljabar, P., Wolz, R., Rueckert, D., 2012. Manifold learning for medical image registration, segmentation, and classification. In: Suzuki, K. (Ed.), Machine Learning in Computer-aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global.

Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. Neural Comput. 9, 1545–1588.

Anchisi, D., Borroni, B., Franceschi, M., Kerrouche, N., Kalbe, E., Beuthien-Beumann, B., Cappa, S., Lenz, O., Ludecke, S., Marcone, A., Mielke, R., Ortelli, P., Padovani, A., Pelati, O., Pupi, A., Scarpini, E., Weisenbach, S., Herholz, K., Salmon, E., Holthoff, V., Sorbi, S., Fazio, F., Perani, D., 2005. Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. Arch. Neurol. 62, 1728–1733.

Andreasen, N., Minthon, L., Davidsson, P., Vanmechelen, E., Vanderstichele, H., Winblad, B., Blennow, K., 2001. Evaluation for CSF-tau and CSF-Abeta42 as diagnostic markers for Alzheimer disease in clinical practice. Arch. Neurol. 58, 373–379.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15, 1373–1396.

Blennow, K., Hampel, H., 2003. CSF markers for incipient Alzheimer's disease. Lancet Neurol. 2, 605–613.

Braak, H., Braak, E., 1998. Evolution of neuronal changes in the course of Alzheimer's disease. J. Neural Transm.-Supp. 53, 127–140.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.

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

Clarkson, M.J., Ourselin, S., Nielsen, C., Leung, K.K., Barnes, J., Whitwell, J.L., Gunter, J.L., Hill, D.L.G., Weiner, M.W., Jack Jr., C.R., Fox, N.C., the Alzheimer's Disease Neuroimaging Initiative, 2009. Comparison of phantom and registration scaling corrections using the ADNI cohort. Neuroimage 47, 1506–1513.

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

Cox, T.F., Cox, M.A.A., 2001. Multidimensional Scaling. Chapman and Hall.

Criminisi, A., Shotton, J., Konukoglu, E., 2012. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends Comput. Graph. Vis. 7, 81–227.

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

Dawbarn, D., Allen, S.J. (Eds.), 2007. Neurobiology of Alzheimer's Disease, 3rd edition. Oxford University Press, New York.

Freeborough, P.A., Fox, N.C., Kitney, R.I., 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. Comput. Methods Programs Biomed. 53, 15–25.

Gerber, S., Tasdizen, T., Fletcher, T.P., Whitaker, R., the Alzheimer's Disease Neuroimaging Initiative, 2010. Manifold modeling for brain population analysis. Med. Image Anal. 14, 643–653.

Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage 57, 378–390.

Gousias, I.S., Rueckert, D., Heckemann, R.A., Dyet, L.E., Boardman, J.P., Edwards, A.D., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. Neuroimage 40, 672–684.

Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., 2011. Random forest-based manifold learning for classification of imaging data in dementia. Machine Learning in Medical Imaging Workshop at 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'11), pp. 159–166.

Gray, K.R., Wolz, R., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., for the Alzheimer's Disease Neuroimaging Initiative, 2012. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage 60, 221–229.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–247.

Hampel, H., Burger, K., Teipel, S.J., Bokde, A.L., Zetterberg, H., Blennow, K., 2008. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimers Dement. 4, 38–48.

Hansson, O., Zetterberg, H., Buchlave, P., Londos, E., Blennow, K., Minthon, L., 2006. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. Lancet Neurol. 5, 228–234.

Hastie, T., Tibshirani, R., Friedman, J., 2011. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York . (corrected 5th printing of 2nd edition).

Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., the Alzheimer's Disease Neuroimaging Initative, 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51, 221–227.

Heckemann, R.A., Keihaninejad, S., Aljabar, P., Gray, K.R., Nielsen, C., Rueckert, D., Hajnal, J.V., Hammers, A., the Alzheimer's Disease Neuroimaging Initiative, 2011. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. Neuroimage 56, 2024–2037.

Herholz, K., Salmon, E., Perani, D., Baron, J.C., Holthoff, V., Frolich, L., Schonknecht, P., Ito, K., Mielke, R., Kalbe, E., Zundorf, G., Delbeuck, X., Pelati, O., Anchisi, D., Fazio, F., Kerrouche, N., Desgranges, B., Eustache, F., Beuthien-Baumann, B., Menzel, C., Schroder, J., Kato, T., Arahata, Y., Henze, M., Heiss, W.D., 2002. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. Neuroimage 17, 302–316.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G.F., Chung, M.K., Johnson, S.C., the Alzheimer's Disease Neuroimaging Initiative, 2009. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., the Alzheimer's Disease Neuroimaging Initiative, 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55, 574–589.

Ho, T., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844.

Hoffman, E.J., Cutler, P.D., Digby, W.M., Mazziotta, J.C., 1990. 3-D phantom to simulate cerebral blood-flow and metabolic images for PET. IEEE Trans. Nucl. Sci. 37, 616–620.

Hoffman, J.M., Welsh-Bohmer, K.A., Hanson, M., Crain, B., Hulette, C., Earl, N., Coleman, R.E., 2000. FDG PET imaging in patients with pathologically verified dementia. J. Nucl. Med. 41, 1920–1928.

Hope, T.A., Gregson, P.H., Linney, N.C., Schmidt, M.H., Abdolell, M., 2008. Selecting and assessing quantitative early ultrasound texture measures for their association with cerebral palsy. IEEE Trans. Med. Imaging 27, 228–236.

Iglesias, J.E., Liu, C.Y., Thompson, P.M., Tu, Z., 2011. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans. Med. Imaging 30, 1617–1634.

Joshi, A., Koeppe, R.A., Fessler, J.A., 2009. Reducing between scanner differences in multi-center PET studies. Neuroimage 46, 154–159.

Keihaninejad, S., Heckemann, R.A., Gousias, I.S., Hajnal, J.V., Duncan, J.S., Aljabar, P., Rueckert, D., Hammers, A., 2012. Classification and lateralization of temporal lobe epilepsies with and without hippocampal atrophy based on whole-brain automatic MRI segmentation. PLoS One 7, e33096.

Klunk, W.E., Engler, H., Nordberg, A., Wang, Y.M., Blomqvist, G., Holt, D.P., Bergstrom, M., Savitcheva, I., Huang, G.F., Estrada, S., Ausén, B., Debnath, M.L., Barletta, J., Price, J.C., Sandell, J., Lopresti, B.J., Wall, A., Koivisto, P., Antoni, G., Mathis, C.A., Långström, B., 2004. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann. Neurol. 55, 306–319.

Landau, S.M., Harvey, D., Madison, C.M., Reiman, E.M., Foster, N.L., Aisen, P.S., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Jack Jr., C.R., Weiner, M.W., Jagust, W.J., 2010. Comparing predictors of conversion and decline in mild cognitive impairment. Neurology 75, 230–238.

Langbaum, J.B.S., Chen, K., Lee, W., Reschke, C., Bandy, D., Fleisher, A.S., Alexander, G.E., Foster, N.L., Weiner, M.W., Koeppe, R.A., Jagust, W.J., Reiman, E.M., the Alzheimer's Disease Neuroimaging Initiative, 2009. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Neuroimage 45, 1107–1116.

Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S., the Alzheimer's Disease Neuroimaging Initiative, 2011. Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. Neuroimage 55, 1091–1108.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis, 5th edition. Academic Press, London.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack Jr., C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 263–269.

Motter, R., Vigopelfrey, C., Kholodenko, D., Barbour, R., Johnsonwood, K., Galasko, D., Chang, L., Miller, B., Clark, C., Green, R., Olson, D., Southwick, P., Wolfert, R., Munroe, B., Lieberburg, I., Seubert, P., Schenk, D., 1995. Reduction of beta-amyloid peptide(42) in the cerebrospinal fluid of patients with Alzheimer's disease. Ann. Neurol. 38, 643–648.

Patwardhan, M.B., McCrory, D.C., Matchar, D.B., Samsa, G.P., Rutschmann, O.T., 2004. Alzheimer disease: operating characteristics of PET — a meta-analysis. Radiology 231, 73–80.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. J. Intern. Med. 256, 183–194.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308.

Rocca, W.A., Hofman, A., Brayne, C., Breteler, M.M.B., Clarke, M., Copeland, J.R.M., Dartigues, J.F., Engedal, K., Hagnell, O., Heeren, T.J., Jonker, C., Lindesay, J., Lobo, A., Mann, A.H., Mölsä, P.K., Morgan, K., O'Connor, D.W., Droux, A.d.S., Sulkava, R., Kay, D.W.K., Amaducci, L., 1991. Frequency and distribution of Alzheimer's disease in Europe: a collaborative study of 1980–1990 prevalence findings. Ann. Neurol. 30, 381–390.

Roses, A.D., Saunders, A.M., 1997. ApoE, Alzheimer's disease, and recovery from brain stress. Cerebrovasc. Pathol. Alzheimers Dis. 826, 200–212.

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Thompson, P.M., Jack Jr., C.R., Weiner, M.W., the Alzheimer's Disease Neuroimaging Initiative, 2009. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132, 1067–1077.

Selkoe, D.J., 1991. The molecular pathology of Alzheimer's disease. Neuron 6, 487–498.

Shi, T., Horvath, S., 2006. Unsupervised learning with random forest predictors. J. Comput. Graph. Stat. 15, 118–138.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Mod. Pathol. 18, 547–557.

Silverman, D.H.S., Small, G.W., Chang, C.Y., Lu, C.S., de Aburto, M.A.K., Chen, W., Czernin, J., Rapoport, S.I., Pietrini, P., Alexander, G.E., Schapiro, M.B., Jagust, W.J., Hoffman, J.M., Welsh-Bohmer, K.A., Alavi, A., Clark, C.M., Salmon, E., de Leon, M.J., Mielke, R., Cummings, J.L., Kowell, A.P., Gambhir, S.S., Hoh, C.K., Phelps, M.E., 2001. Positron emission tomography in evaluation of dementia — regional brain metabolism and long-term outcome. JAMA 286, 2120–2127.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack Jr., C.R., Kaye, J., Montine, T.J., Park, D.C., Reiman, E.M., Rowe, C.C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M.C., Thies, B., Morrison-Bogorad, M., Wagster, M.V., Phelps, C.H., 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 280–292.

Tapiola, T., Pirttila, T., Mehta, P.D., Alafuzoff, I., Lehtovirta, M., Soininen, H., 2000. Relationship between apoE genotype and CSF beta-amyloid (1–42) and tau in patients with probably and definite Alzheimer's disease. Neurobiol. Aging 21, 735–740.

Tenenbaum, J., de Silva, V., Langford, J., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.

Torgerson, W.S., 1952. Multidimensional scaling: I. Theory and method. Psychometrika 17, 401–419.

Trojanowski, J.Q., Vandeerstichele, H., Korecka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W.Z., Weiner, M.W., Jack Jr., C.R., Jagust, W., Toga, A.W., Lee, V.M.Y., Shaw, L.M., 2010. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. Alzheimers Dement. 6, 230–238.

van de Pol, L.A., van der Flier, W.M., Korf, E.S.C., Fox, N.C., Barkhof, F., Scheltens, P., 2007. Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. Neurology 69, 1491–1497.

Vandermeeren, M., Mercken, M., Vanmechelen, E., Six, J., Vandevoorde, A., Martin, J.J., Cras, P., 1993. Detection of tau proteins in normal and Alzheimer's disease cerebrospinal fluid with a sensitive sandwich enzyme-linked immunosorbent assay. J. Neurochem. 61, 1828–1834.

Walhovd, K.B., Fjell, A.M., Dale, A.M., McEvoy, L.K., Brewer, J., Karow, D.S., Salmon, D.P., Fennema-Notestine, C., 2010. Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiol. Aging 31, 1107–1121.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Siuciak, J.A., Soares, H., Toga, A.W., Trojanowski, J.Q., the Alzheimer's Disease Neuroimaging Initiative, 2012. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement. 8, S1–S68.

Yakushev, I., Hammers, A., Fellgiebel, A., Schmidtmann, I., Scheurich, A., Buchholz, H.G., Peters, J., Bartenstein, P., Lieb, K., Schreckenberger, M., 2009. SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. Neuroimage 44, 43–50.

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

Zhang, D., Shen, D., the Alzheimer's Disease Neuroimaging Initiative, 2012. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS One 7, e33182.