NeuroImage 60 (2012) 221–229

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease

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

[a] Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK
[b] Fondation Neurodis, CERMEP-Imagerie du Vivant, Lyon, France
[c] Division of Neuroscience and Mental Health, Faculty of Medicine, Imperial College London, UK

ARTICLE INFO

Article history:
Received 31 July 2011
Revised 15 December 2011
Accepted 24 December 2011
Available online 6 January 2012

Keywords:
Alzheimer's disease
Mild cognitive impairment
Classification
Longitudinal analysis
[18F]fluorodeoxyglucose positron emission tomography
Image segmentation

ABSTRACT

Imaging biomarkers for Alzheimer's disease are desirable for improved diagnosis and monitoring, as well as drug discovery. Automated image-based classification of individual patients could provide valuable diagnostic support for clinicians, when considered alongside cognitive assessment scores. We investigate the value of combining cross-sectional and longitudinal multi-region FDG-PET information for classification, using clinical and imaging data from the Alzheimer's Disease Neuroimaging Initiative. Whole-brain segmentations into 83 anatomically defined regions were automatically generated for baseline and 12-month FDG-PET images. Regional signal intensities were extracted at each timepoint, as well as changes in signal intensity over the follow-up period. Features were provided to a support vector machine classifier. By combining 12-month signal intensities and changes over 12 months, we achieve significantly increased classification performance compared with using any of the three feature sets independently. Based on this combined feature set, we report classification accuracies of 88% between patients with Alzheimer's disease and elderly healthy controls, and 65% between patients with stable mild cognitive impairment and those who subsequently progressed to Alzheimer's disease. We demonstrate that information extracted from serial FDG-PET through regional analysis can be used to achieve state-of-the-art classification of diagnostic groups in a realistic multi-centre setting. This finding may be usefully applied in the diagnosis of Alzheimer's disease, predicting disease course in individuals with mild cognitive impairment, and in the selection of participants for clinical trials.

Introduction

Alzheimer's disease (AD) is the most common cause of dementia in the elderly, with a worldwide prevalence of 26.6 million reported in 2006, which is expected to rise above 100 million by 2050 (Brookmeyer et al., 2007). Drug development is a major research focus, but at present no treatment is able to either reduce the risk of developing AD, or delay its onset and progression. Any disease-modifying or causal therapy would likely be of greatest benefit to pre-symptomatic patients, and those at increased risk of developing AD. Patients with amnestic mild cognitive impairment (MCI) are hence of particular interest for clinical trials.

Consensus diagnostic criteria for established AD are up to 90% accurate when validated against neuropathological gold standards (Ranginwala et al., 2008). There are, however, several significant challenges to be addressed. These include pre-symptomatic diagnosis, differential diagnosis and the assessment and prediction of progression. Population stratification is also important, to allow recruitment of appropriate participants for clinical trials, and targeting of patients for whom newly developed treatments may be most effective. Recently published revisions to the consensus criteria aim to incorporate advances in AD research, such as the diagnostic and prognostic value of biochemical and neuroimaging biomarkers (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011). The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a valuable resource for related investigations, providing longitudinal clinical and imaging data from patients with AD and MCI, as well as healthy controls (HC).

Functional imaging with FDG-PET is one of several neuroimaging modalities of interest in AD. Numerous studies (for example, Langbaum et al., 2009; Mosconi et al., 2005, 2008; Nestor et al., 2003) have shown that both MCI and AD are associated with significant reductions in the cerebral metabolic rate of glucose in brain regions preferentially affected

* 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://www.loni.ucla.edu/ADNI). 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 is available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.12.071

by the disease. AD patients typically display reductions of greater magnitude and spatial extent. Reduced metabolic activity in AD patients can predict both their cognitive decline and histopathological diagnosis (Hoffman et al., 2000; Minoshima et al., 2001; Silverman et al., 2001), and in MCI patients it can predict their conversion to AD (Anchisi et al., 2005; Mosconi et al., 2004). Serial FDG-PET scans over ten years can identify declining hippocampal metabolism as healthy individuals progress to AD (de Leon et al., 2001). FDG-PET is mentioned in the revised AD diagnostic criteria (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011) as a potentially useful tool for early diagnosis and monitoring of disease progression. However, as with other neuroimaging and biochemical biomarkers, its use is recommended for research, rather than standard clinical practice.

Automated image-based classification of individual patients could provide valuable diagnostic support for clinicians, when considered alongside cognitive assessment scores. The ADNI study provides an ideal dataset for classification, since it approximates a clinical population due to its large size and diversity. Several recent studies have performed image-based classification using cross-sectional ADNI FDG-PET data. Hinrichs et al. (2009) use spatially augmented linear programme boosting, based on voxel-wise features, to achieve a classification accuracy of 84% between AD patients and HC. Haense et al. (2009) apply a previously validated method (Herholz et al., 2002) in which a global measure of image abnormality is provided by the sum of abnormal t-values in predefined areas typically affected by AD. They report 83% sensitivity and 78% specificity between AD patients and HC following the application of a preset t-sum threshold. Salas-Gonzalez et al. (2010) apply a linear support vector machine (SVM) to voxel-wise data which have undergone feature selection and dimensionality reduction. They achieve accuracies of 87% between AD patients and HC, and 83% between MCI patients and HC, using a two-fold cross-validation strategy.

There are many more classification studies based on ADNI structural MR imaging data. For example, a recent study (Cuingnet et al., 2011) compares ten high-dimensional classification methods applied to 509 baseline ADNI 1.5 T MR images. Two methods use only the hippocampal shape or volume, whilst the remainder are whole-brain approaches, which use either cortical thickness measures, or voxel-wise tissue class probabilities for grey matter, white matter and cerebrospinal fluid (CSF). High accuracies in distinguishing AD patients from HC (up to 81% sensitivity and 95% specificity) are reported for whole-brain approaches. Four of the ten methods were able to distinguish MCI patients who later progressed to AD (pMCI) from those who remained stable (sMCI) over 18 months slightly more accurately than a random classifier, although not significantly ($p > 0.05$) so. For example, one of the methods based on hippocampal volume achieved 62% sensitivity and 69% specificity.

There is increasing interest in using multi-modality imaging and non-imaging data for classification. For example, Zhang et al. (2011) apply a kernel combination approach to cross-sectional FDG-PET and MR imaging data and CSF biomarker measures. They report classification accuracies of 93% between AD patients and HC, and 76% between MCI patients and HC, when using all three modalities in combination. These results are superior to those obtained when using any one modality independently. Hinrichs et al. (2011) have also investigated the application of kernel combination methods, but to both cross-sectional and longitudinal FDG-PET and MR imaging data, as well as CSF biomarker measures, neuropsychological status examination scores, and APOE genotype information. They, too, report that the use of multi-modality data leads to superior classification performance compared with that based on any individual modality. Their study data included two FDG-PET and MR images for each subject, taken approximately 24 months apart. They observed that longitudinal analysis of the FDG-PET images (either voxel-wise temporal difference or temporal ratio) performed relatively poorly in distinguishing AD patients from HC, compared with the raw FDG-PET signal intensities at either timepoint. They suggest that two-year changes in FDG-PET signal intensity alone are not sufficient to identify AD with high accuracy, and these longitudinal data were therefore not incorporated into their multi-modality classification experiments.

In contrast, Chen et al. (2010) report highly significant group differences between AD or MCI patients and HC in their longitudinal analysis of 12-month metabolic declines in ADNI subjects. This suggests that whilst longitudinal FDG-PET data alone may not be sufficient for classification, they may provide valuable complementary information which can enhance the results achievable using cross-sectional FDG-PET. We therefore investigate the value of combining cross-sectional and longitudinal FDG-PET information for classification. We extract regional features from baseline and 12-month follow-up FDG-PET images, and investigate their combined use for image-based classification of the ADNI participants. We present classification results for four clinically relevant pairs of diagnostic groups (AD/HC, pMCI/HC, AD/sMCI, pMCI/sMCI), and also identify the regional features which best separate these groups.

Materials and methods

Imaging data

Data used in the preparation of this article were obtained from the ADNI database (http://www.loni.ucla.edu/ADNI). 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. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. 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, M.D., 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, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research — approximately 200 cognitively normal older individuals to be followed for three years, 400 people with MCI to be followed for three years, and 200 people with early AD to be followed for two years. Further up-to-date information, including detailed eligibility criteria, is available on the ADNI information website (http://www.adni-info.org).

Baseline and 12-month follow-up FDG-PET and 1.5 T MR images were available to download for 321 ADNI participants. We excluded from our analysis any subjects for whom one or both of the FDG-PET images were acquired using the Siemens HRRT or BioGraph HiRez scanners ($n = 78$), due to differences in the observed 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). We also excluded a small number of subjects ($n = 10$) whose images could not be processed as required, either because of missing timeframe information in the FDG-PET image headers, or incorrect positioning of subjects in the PET scanner, such that part of the brain was outside the field of view. The MCI subjects were divided into pMCI and sMCI based on changes in clinical status occurring over 19 ± 10 (range 6–48) months. Subjects whose diagnosis did not clearly fall into one of the four clinical categories (AD, pMCI, sMCI, HC) were additionally excluded ($n = 12$). Of these 12, five progressed from HC to MCI, five reverted from MCI to HC, and two oscillated between MCI and HC. ADNI subject identifiers for the 22 excluded subjects are provided as Supplementary data.

We use imaging data from 221 participants (50 AD, 53 pMCI, 64 sMCI, 54 HC), whose groupwise characteristics are provided in Table 1. The mean age at baseline (75.7 ± 6.3 years) and mean time between baseline and 12-month FDG-PET scans (11.6 ± 0.9 months) do not vary significantly ($p > 0.01$) on t-test between the clinical groups.

Image processing

Our analysis is performed in native MRI space. Baseline anatomical segmentations were automatically generated in the space of the baseline MRI, and follow-up segmentations were produced by nonlinear registration to the space of the 12-month MRI. FDG-PET images were co-registered with their corresponding MR images. An independently derived reference cluster required for FDG-PET image normalisation was provided in MNI space (Mazziotta et al., 1995), and transformed into the baseline and 12-month MRI space of each subject. An overview of the image processing pipeline is illustrated in Fig. 1, and full details of each step are provided in the sections that follow.

ADNI FDG-PET acquisition

The FDG-PET images had been acquired using Siemens, GE and Philips PET scanners according to one of three standard protocols (30–60 minute dynamic, 30–60 minute static, 0–60 minute dynamic) following the intravenous injection of 185 ± 19 MBq of FDG. Data were corrected for both scatter and measured attenuation, which was determined using the CT scan for PET/CT scanners, and a transmission scan with ^{68}Ge or ^{137}Cs rotating rod sources for PET-only scanners. Images were reconstructed using scanner-specific algorithms, and sent to the University of Michigan, where they were reviewed for artefacts, de-identified, and transmitted to the Laboratory of NeuroImaging (LONI) for storage. Further details are available in the ADNI PET technical procedures manual (ADNI PET Core, 2005).

FDG-PET image pre-processing

The 221 baseline and 12-month FDG-PET scans were downloaded from the LONI Image Data Archive in their original DICOM or ECAT format. They were converted to NIfTI using (X)MedCon (http://xmedcon.sourceforge.net), with care taken to preserve negative values and correctly apply any quantification factors. Each image was examined for major artefacts, and its orientation adjusted if necessary. The 30–60 minute dynamic scans were corrected for patient motion using tools from the Image Registration Toolkit (IRTK; http://www.doc.ic.ac.uk/~dr/software) to register each of the subsequent frames rigidly to the image's first frame. The resulting co-registered frames were averaged to produce a single 30–60 minute static image. For the 0–60 minute dynamic scans, the final six 5-minute frames were extracted, and concatenated into a static image in the same way.

ADNI MRI acquisition and pre-processing

Pre-processed versions of the 221 baseline and 12-month T1-weighted 1.5 T MRI scans were downloaded from the LONI Image Data Archive in NIfTI format. These had been acquired according to a standard protocol (Jack et al., 2008) involving two scans per subject that were based on a 3-D MPRAGE imaging sequence and acquired using Siemens, GE and Philips MRI scanners. Further details are available in the ADNI MRI technical procedures manual (ADNI MRI Core, 2005). Of the two images acquired per subject and timepoint, the ADNI quality assurance team selected the better image for pre-processing, based on the presence and severity of common image artefacts, as well as other criteria. Pre-processing involved the application of a scanner-specific correction for gradient non-linearity distortion (Gradwarp; Jovicich et al., 2006), followed by a correction for image intensity non-uniformity (B1; Jack et al., 2008), and finally a histogram peak sharpening algorithm for bias field correction (N3; Sled et al., 1998). Only the N3 pre-processing step was necessary for images acquired on Philips scanners, since B1 correction was already implemented, and their gradient systems tended to be linear (Jack et al., 2008).

Co-registration of FDG-PET with MRI

For each subject and timepoint, the pre-processed FDG-PET image was co-registered with the corresponding pre-processed MR image, and re-sampled to the higher resolution of the MRI. Tools from IRTK were used to perform rigid and then affine registration, using normalised mutual information as the similarity criterion (Studholme et al., 1999), and the affine transformation parameters were applied to the FDG-PET image using a linear interpolation. An affine transformation was preferred over a rigid one because it can account for any scaling or voxel size errors which may remain after phantom correction of the MRI (Clarkson et al., 2009).

Baseline MRI anatomical segmentation

Automatic whole-brain segmentations into 83 anatomical regions were prepared in the native space of each baseline MRI using multi-atlas propagation with enhanced registration (MAPER), an approach that has been previously described and validated for use in subjects with AD, and age-matched HC (Heckemann et al., 2010). The segmentations are available to download through the ADNI website, and full details of the segmentation 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).

Individual tissue probability maps for CSF, grey matter and white matter were obtained using FSL FAST (http://www.fmrib.ox.ac.uk/fsl). For FDG-PET image analysis, the grey matter portion within each cortical label is of relevance. Masked segmentations were therefore employed, in which all regions except ventricles, central structures, cerebellum and brainstem had been masked with a grey matter label, and the lateral ventricles with a CSF label.

12-month MRI anatomical segmentation

To obtain similar whole-brain anatomical segmentations for the follow-up images, we propagated each baseline segmentation to the space of the corresponding 12-month MR image using nonrigid registration. The intracranial portion of the 12-month MRI was first determined by rigid propagation of the baseline intracranial mask that had been used for brain extraction during the MAPER segmentation procedure. The baseline intracranial masks were derived, as described in Heckemann et al. (2011), from binary masks covering intracranial white and grey matter. These binary masks had been generated using MIDAS, a semi-automatic procedure described elsewhere (Freeborough et al., 1997; Leung et al., 2011). The rigidly aligned intracranial-masked baseline and 12-month MRI were then affinely aligned, again to account for possible scaling or voxel size errors, followed by a series of nonrigid registrations. The nonrigid registration

Table 1
Clinical and demographic information for the study population. For each clinical group, the total number of subjects (N) and number of females (F) are shown, along with the baseline clinical dementia rating (CDR), average baseline mini-mental state examination (MMSE) score, and average change in MMSE score over the 12-month follow-up period (Δ_{MMSE}).

	N(F)	CDR (%)			MMSE	Δ_{MMSE}
		0	0.5	1	(mean ± std. dev.)	(mean ± std. dev.)
AD	50 (20)	0	36	64	23.5 ± 2.0	−2.76 ± 3.96
pMCI	53 (20)	0	100	0	26.7 ± 1.7	−1.79 ± 2.57
sMCI	64 (18)	0	100	0	27.5 ± 1.7	0.05 ± 1.79
HC	54 (18)	100	0	0	28.9 ± 1.2	0.30 ± 1.47

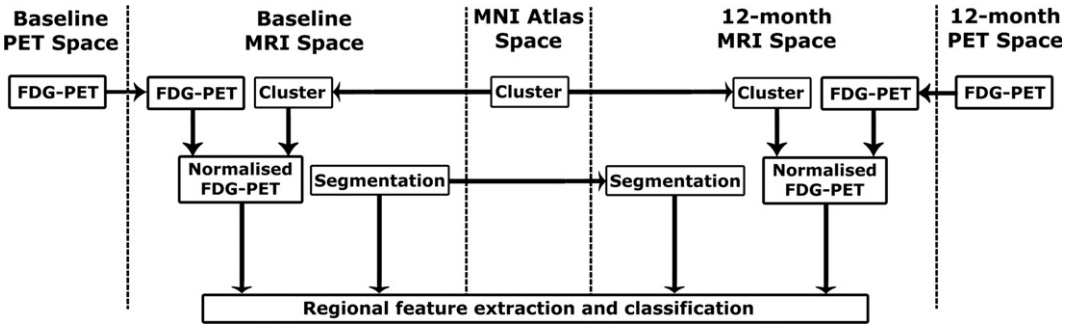

Fig. 1. Image processing pipeline, illustrating the images required for regional feature extraction from the baseline and 12-month FDG-PET images. Horizontal arrows indicate image registration and re-slicing steps. Vertical arrows indicate images used for feature extraction.

used was a free-form deformation, with a flexibility defined by the spacing of a lattice of control points (Rueckert et al., 1999; Schnabel et al., 2001). Nonrigid registration was performed using control point spacings of 10, 5, and 2.5 mm. The unmasked baseline anatomical segmentation was nonrigidly propagated to 12-month MRI space using nearest neighbour interpolation. Individual tissue probability maps for CSF, grey matter and white matter were obtained for the 12-month MRI using FSL FAST, and the segmentation masked using the same procedure as for the baseline.

FDG-PET normalisation

FDG-PET image normalisation is often performed relative to the cerebral global mean. However, due to the nature of the disease process, both MCI and AD patients have a lower glucose metabolic rate than HC across the whole brain. Normalisation to the cerebral global mean therefore artificially scales up values from patients, whilst scaling down those from HC, resulting in under-estimation of the relative hypometabolism in patients compared to HC (Yakushev et al., 2008). In addition, such normalisation results in areas of apparent hypermetabolism being observed in patients compared to HC in regions of the brain that are relatively preserved in AD, including the cerebellum, brainstem, basal ganglia, and sensorimotor cortex (Herholz et al., 2002). Recent work suggests that improved group discrimination can be achieved by using the signal intensity in these relatively preserved regions of the brain for normalisation, rather than the cerebral global mean value (Borghammer et al., 2009; Yakushev et al., 2009). Our analysis makes use of this "reference cluster" normalisation method.

We obtained a MNI space image of the reference cluster used in Yakushev et al. (2009) from the author. Using the "Segment" module of SPM5 (http://www.fil.ion.ucl.ac.uk/spm), each baseline and 12-month MRI was linearly and non-linearly deformed (Ashburner and Friston, 2005) to the MNI template. The inverse transformation was used to map the MNI-space cluster into the native MRI space of each subject and timepoint using trilinear interpolation. The cluster was also re-sampled to the higher resolution of the MRI.

Regional feature extraction and classification

Each of the MRI-space FDG-PET images was overlaid with its corresponding masked anatomical segmentation. The FDG-PET signal intensity per mm^3 was determined for each of the 83 anatomically defined regions. Global variations in the cerebral metabolic rate of glucose between subjects were accounted for by normalisation to the signal intensity per mm^3 in the reference cluster. Typical examples of the images required for regional feature extraction are shown in Fig. 2. Normalised regional signal intensities were thus extracted from both the baseline and 12-month FDG-PET images. Additionally, the regional changes in FDG-PET signal intensity over the 12-month follow-up period were determined. Each subject therefore had 249 regional features available for use in classification experiments.

We performed classification using a SVM classifier with LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm), an integrated software for support vector classification (Chang and Lin, 2011). Robust estimates of classifier performance were obtained via a repeated random sampling approach, assessing the classification rates between four clinically relevant pairs of diagnostic groups (AD/HC, pMCI/HC, AD/sMCI, pMCI/sMCI). The mean classification accuracy, sensitivity, specificity and balanced error rate for pairs of groups were evaluated over 1000 runs, in which 75% of the subjects were randomly selected for training, with the remaining 25% used as test data. We tested five regional feature sets: baseline signal intensities, 12-month signal intensities, relative changes in signal intensity over 12 months, baseline signal intensities concatenated with 12-month changes, and 12-month signal intensities concatenated with 12-month changes. For each clinical group pair, unpaired *t*-tests between the distributions of classification results obtained from the 1000 leave-25%-out runs were performed to assess the significance of differences in performance between the five feature sets. Since we accounted for the relative sizes of the clinical groups when sampling the training sets and training the SVM classifier, the sensitivity and specificity values obtained were fairly well balanced, and the balanced error rate was very similar to the total accuracy. We therefore selected total accuracy as an overall performance metric, because this allowed for more direct comparisons with other published works which quote classification performance in terms of accuracy.

In addition, to allow a better assessment of the statistical significance of our results, we performed permutation testing for all classification experiments. For each pair of clinical groups, the diagnostic labels were randomly permuted, the data divided into training and test sets, and the SVM classifier trained and tested as described above. This process was repeated 1000 times per clinical group pair. Permutation tests thus provide a distribution of classification accuracies under the null hypothesis, that the classifier cannot accurately predict the clinical labels from the data provided. To assess whether our observed classification accuracy for the clinical group pair was significantly different from chance, we therefore performed an unpaired *t*-test between the distribution of observed accuracies, and the distribution obtained from permutation testing.

A two-class SVM aims to construct a hyperplane that maximises the margin, which is the distance between the closest points on either side of the boundary, known as the support vectors. For a set of training data $(\mathbf{X}, \mathbf{t}) = \{(\mathbf{x_i}, t_i)\}_{i=1}^{N}$, where each subject has a feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_D)$, and a class label $t \in \{-1, 1\}$, a hyperplane separating the classes may be written $y(\mathbf{x}) = \mathbf{w^T x} - b = 0$. Since the data were unlikely to be linearly separable, we applied a soft-margin formulation of the SVM (LIBSVM "C-SVC"), in which the trade-off between maximising the margin and minimising the training error is controlled by the penalty parameter *C*. Additionally, the data were transformed into a

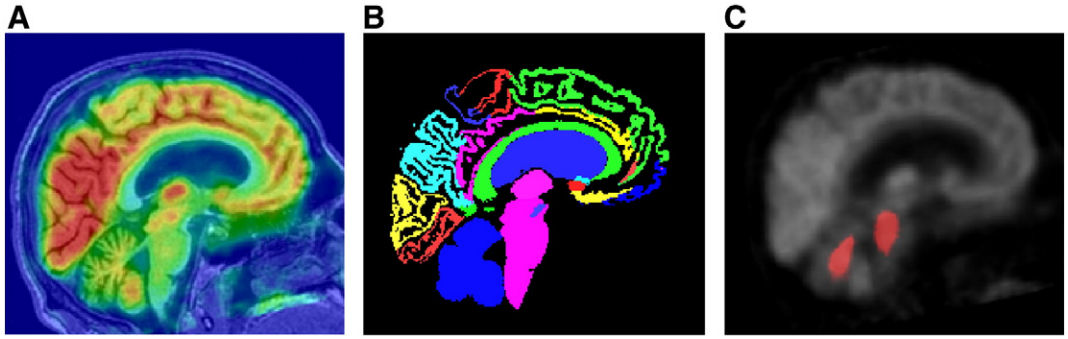

Fig. 2. Typical examples of the images required for regional feature extraction from the baseline images of a HC subject. From left to right: baseline MRI overlaid with baseline FDG-PET; masked anatomical segmentation; baseline FDG-PET overlaid with normalisation cluster. The regional colour map for the segmentation is as used in Gousias et al. (2008).

higher-dimensional space using the nonlinear function $\boldsymbol{\phi}$. A boundary that is nonlinear in the original feature space may be better approximated by the linear boundary $y(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\phi(\mathbf{x}) - b = 0$ in the transformed space. This requires the solution of the optimisation problem (Cortes and Vapnik, 1995).

$$\min_{\mathbf{w},\xi,b} \left\{ \frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} + C\sum_{i=1}^{N}\xi_i \right\}$$
$$\text{subject to } t_i\left(\mathbf{w}^{\mathbf{T}}\phi(x_i) - b\right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0,$$

where $\boldsymbol{\xi}$ are slack variables which model the allowable degree of misclassification. The nonlinear mapping was performed using a radial basis function kernel $k\left(\mathbf{x_i}, \mathbf{x_j}\right) \equiv \phi(\mathbf{x_i})^{\mathbf{T}}\phi\left(\mathbf{x_j}\right) = \exp\left(-\gamma\|\mathbf{x_i} - \mathbf{x_j}^2\|\right)$ of width $\gamma > 0$. As part of the training process, it was necessary to optimise the parameters C and γ. This was achieved by performing a grid-search using five-fold cross-validation, such that the (C, γ) pair resulting in the highest cross-validation accuracy was selected. The SVM classifier was then trained using the full set of training data, before having its performance assessed on the test data.

Results

Classification experiments

Two sets of cross-sectional features had been extracted for each subject (regional signal intensities at each of the two imaging timepoints). For all four clinical group pairs, highly significant ($p < 0.001$) increases in classification accuracy were achieved when using 12-month signal intensities compared with using baseline signal intensities. Longitudinal features had also been evaluated as the relative changes in signal intensity over the 12-month follow-up period. For each of the four pairs of clinical groups, classification based on the longitudinal information alone had significantly ($p < 0.05$) lower accuracy compared with using either of the two cross-sectional feature sets.

We also assessed classification performance based on two feature sets which combined the cross-sectional and longitudinal information. These were formed by concatenating the longitudinal change features with the signal intensities at either imaging timepoint. For each clinical group pair, highly significant ($p < 0.001$) increases in classification accuracy were achieved when combining longitudinal information with 12-month data, compared with its combination with baseline data. In addition, classification based on the combination of longitudinal data with 12-month signal intensities was significantly ($p < 0.05$) improved compared with using 12-month signal intensities alone.

The above results are illustrated as boxplots in Fig. 3, and numerical results are provided in Table 2 for the two cross-sectional feature sets and the best-performing combined feature set (longitudinal change concatenated with 12-month signal intensities). Receiver operating characteristic (ROC) curves for classification based on this combined feature set are displayed in Fig. 4, along with the area under each curve (AUC), which provides an overall measure of classifier performance. All classification accuracies were significantly different from chance, as assessed by permutation testing.

To demonstrate that classification was truly based on disease-specific imaging information, rather than the intrinsic age and gender information captured in the images, we additionally performed classification after accounting for these effects. A linear regression step was incorporated into the classification procedure for every clinical group pair such that, for each of the 1000 repetitions, a regression model was estimated from the training data, and the SVM trained on the residuals. The regression model was then applied to the test data, and the SVM tested on the resulting residuals. Regression had no significant effect on the classification accuracy for the majority of experiments. Using the best performing combined feature set (longitudinal change concatenated with 12-month signal intensities), accuracies after linear regression for gender and age at scan were not significantly different for AD/HC, pMCI/HC, or AD/sMCI. However, the mean accuracy improved from 63% to 64% for pMCI/sMCI.

Regional features

We performed t-tests between pairs of clinical groups to identify the regional features which give significant ($p < 0.01$, uncorrected for multiple comparisons) differences between diagnostic groups. We considered both sets of cross-sectional features (baseline and 12-month regional signal intensities), as well as the regional changes in signal intensity over the 12-month follow-up period.

The overwhelming majority of regions differed significantly between AD patients and HC for both baseline intensities (65/83 regions), and 12-month intensities (73/83 regions). For the 12-month data, as well as more regions reaching significance, significance levels were higher than for the baseline data. Far fewer regions reached significance for the change features (26/83 regions), and significance levels were lower than for either of the cross-sectional feature sets. These results are illustrated in Fig. 5, and similar patterns were consistently observed across the remaining three clinical group pairs (pMCI/HC, AD/sMCI, pMCI/sMCI). For these pairs of groups, fewer regions reached significance than between AD patients and HC, and at reduced significance levels. The fewest significant regions, and lowest significance levels were found between pMCI and sMCI patients.

For both sets of cross-sectional features, the five most significantly different regions between AD patients and HC were the bilateral hippocampus, left parietal lobe, left posterior temporal lobe, and right posterior cingulate gyrus. However, only one of these regions (right hippocampus) was amongst the five most significantly different regions for the change features, along with the right amygdala, right middle and

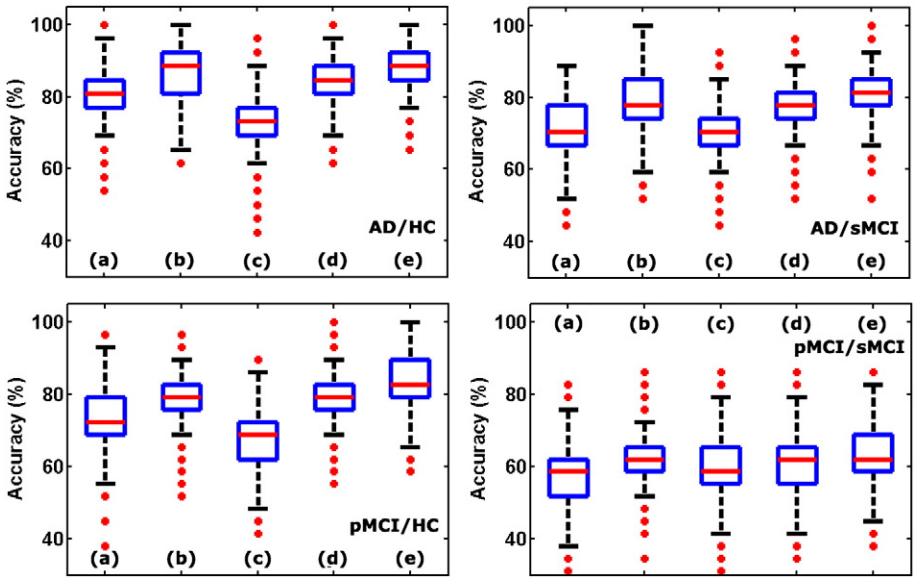

Fig. 3. Classification accuracies for the four clinical group pairs based on the five feature sets studied. From left to right for each boxplot: (a) baseline signal intensities, (b) 12-month signal intensities, (c) change over 12 months, (d) combined baseline intensities and change, and (e) combined 12-month intensities and change. In each boxplot, the central red line represents the median, the edges of the blue box represent the 25th and 75th percentiles, and the black whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually in red for points lying outside of the range $\pm 2.7\sigma$.

inferior temporal gyri, right posterior part of the superior temporal gyrus, and right posterior temporal lobe. For the remaining three group pairs, the five most significantly different regions for each of the three feature sets contained some combination of the regions identified between AD patients and HC, with the parahippocampal gyrus also identified in some cases.

Interestingly, the amygdala was consistently identified amongst the five most significantly different regions for the change features, but not for either of the cross-sectional feature sets. In fact, it was the only region reaching significance for the change features between pMCI and sMCI patients.

Discussion and conclusion

We demonstrate that a combination of cross-sectional and longitudinal FDG-PET information results in classification performance that is in line with the current state-of-the-art. For the most commonly reported classification task of separating AD patients from HC, our accuracy of 88% is comparable with other recent classification results based on multi-modality imaging and non-imaging data (Hinrichs et al., 2011; Zhang et al., 2011), and also with the results of high-dimensional pattern recognition methods applied to cross-sectional MR imaging data (Chupin et al., 2009; Cuingnet et al., 2011). Classification results may well be converging on a "glass ceiling" for this task, since diagnostic consensus criteria themselves have an accuracy of around 90% (Ranginwala et al., 2008). For FDG-PET in particular, it is also important to consider the further confounding factor that approximately 10% of the ADNI AD patients have a pattern of glucose metabolism that is more consistent with frontotemporal dementia (Jagust et al., 2010; Thiele et al., 2009).

We additionally attempt the less commonly reported classification task of separating pMCI from sMCI patients. Our accuracy of 65% is encouraging compared with the most directly comparable studies based on MR imaging data (Cuingnet et al., 2011; Wolz et al., 2010). It has been reported that progression from MCI to AD occurs at a rate of 10–15% per year (Petersen, 1999), with up to 80% of MCI patients developing AD over a six year period (Petersen, 2004). To properly assess the utility of any classification method in predicting such progression, longer clinical follow-up is therefore required than is currently available for the ADNI participants.

Table 2
Classification accuracy (acc), sensitivity (sens), specificity (spec), and balanced error rate (BER) expressed as mean (standard deviation) over the 1000 leave-25%-out runs. Results are provided for baseline signal intensities, 12-month signal intensities, and the feature set combining relative changes with 12-month signal intensities. The effects of gender and age at scan are not regressed out for these data.

		AD/HC	pMCI/HC	AD/sMCI	pMCI/sMCI
Baseline	Acc (%)	80.9 (6.7)	70.7 (7.6)	72.7 (7.8)	58.4 (7.9)
	Sens (%)	79.6 (10.6)	70.5 (13.0)	65.6 (13.2)	51.5 (13.1)
	Spec (%)	82.3 (11.1)	71.0 (12.3)	78.5 (10.8)	64.8 (13.4)
	BER (%)	80.9 (6.7)	70.8 (7.6)	72.0 (7.9)	58.1 (7.9)
12-month	Acc (%)	86.1 (6.3)	79.2 (7.3)	79.3 (6.7)	62.3 (7.8)
	Sens (%)	81.2 (10.3)	77.2 (11.9)	77.9 (11.2)	53.2 (13.1)
	Spec (%)	91.0 (8.2)	81.4 (11.0)	80.5 (9.3)	70.8 (12.4)
	BER (%)	86.1 (6.3)	79.3 (7.3)	79.2 (6.8)	62.0 (7.8)
Combined	Acc (%)	88.4 (6.2)	81.3 (6.8)	83.5 (7.1)	63.1 (8.1)
	Sens (%)	83.2 (10.4)	79.8 (11.1)	79.9 (11.5)	52.2 (13.5)
	Spec (%)	93.6 (7.4)	82.9 (10.7)	86.4 (8.9)	73.2 (12.3)
	BER (%)	88.4 (6.2)	81.4 (6.8)	83.2 (7.2)	62.7 (8.1)

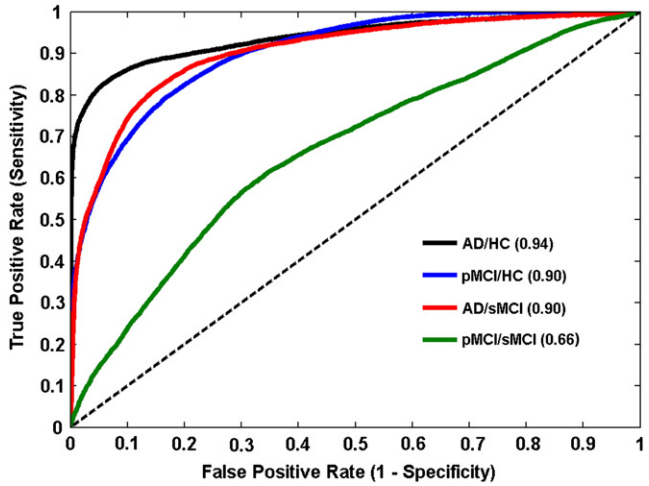

Fig. 4. ROC curves for the combined feature set of relative changes concatenated with 12-month signal intensities. AUC values for each clinical group pair are provided in brackets.

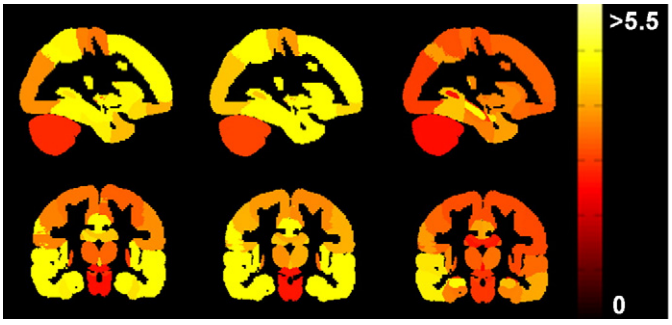

Fig. 5. Regional *t*-values for comparisons between AD patients (n = 50) and HC (n = 54) superimposed onto sagittal (top row) and coronal (bottom row) slices of a maximum probability brain atlas, which has been masked according to the same procedure as the anatomical segmentations. The feature sets tested are, from left to right: baseline signal intensities; 12-month signal intensities; changes in signal intensity over 12 months. To allow all three feature sets to be visualised using the same colour scale, so that their spatial patterns may be compared, all *t*-values greater than 5.5 have been scaled to the maximum value.

To verify that the regional features used for classification made biological sense, we performed *t*-tests between clinical groups to assess which regions gave statistically significant group differences. Although a direct visualisation of the SVM weight vector would be desirable, because of the nonlinearity of the kernel used, it was not possible to map the weight vectors learned in the transformed feature space back to the original feature space in a meaningful way. We therefore explored univariate changes using *t*-tests for the purposes of visualisation. When considering the cross-sectional data, regional *t*-values between AD patients and HC indicated significant differences across most of the brain. This finding is consistent with the voxel-wise *t*-tests reported in Yakushev et al. (2009). The most significantly different regions between groups included those known to be affected in AD for all three feature sets, consistent with previous voxel-wise *t*-tests performed on the ADNI FDG-PET data (Chen et al., 2010; Langbaum et al., 2009).

Similarly to Hinrichs et al. (2011), we found that the percentage change in signal intensity over 12 months alone does not provide particularly impressive classification performance between AD patients and HC (74% accuracy). Although the longitudinal data alone appear insufficient for matching state-of-the-art classification performance, our results demonstrate that they can provide some complementary information which can enhance classification when used in conjunction with cross-sectional features. This suggestion is supported by our *t*-test results, which show that the pattern of regional significances differs between cross-sectional and longitudinal data. For example, the amygdala is identified amongst the best five features for group discrimination only for the longitudinal data. The two cross-sectional feature sets, on the other hand, have similar patterns of regional significances, although improved group discrimination may be achieved with the 12-month data.

We additionally performed all classification experiments after accounting for the effects of age and gender by linear regression. The lack of significant effect on accuracy observed in the majority of cases indicates that the classification results were truly based on disease-specific imaging information, rather than the intrinsic age and gender information also captured in the images. The significant improvement observed between pMCI and sMCI patients is in agreement with our previous findings (Gray et al., 2011b, 2011c). In our previous work, we had performed a global regression, whereby the coefficients for age and gender were estimated using all the available data from all four of the diagnostic groups. We have since demonstrated that there is little appreciable difference between the effect of the two regression approaches.

The aim of this work has not been to introduce a novel classification approach, but instead to use a readily available SVM classifier and simple feature combination approach (direct concatenation) to demonstrate the utility of longitudinal FDG-PET information for improving classification amongst four clinically relevant pairs of diagnostic groups. Having established that the longitudinal features can indeed enhance the results achieved using cross-sectional data alone, it may be beneficial to investigate the application of kernel combination methods, which are reported to be superior to simple concatenation for combining feature sets (Zhang et al., 2011). Additionally, the possibility of multi-class classification could be investigated, for example, using the LIBSVM "one-against-one" multi-class strategy (Chang and Lin, 2011).

An important consideration of the described regional FDG-PET analysis approach is its requirement for MR imaging data. Structural imaging, either with MRI or CT, is routinely used in clinical practice to identify brain lesions that could lead to a clinical picture mimicking a diagnosis of AD. Both MRI and FDG-PET are mentioned in the revised AD diagnostic criteria (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011) as providing potentially useful biomarkers, and the recent development of hybrid MRI-PET technology means that the simultaneous acquisition of both modalities could become a practical solution for dementia imaging in the future. For example, one such system has been approved for use in clinical practice in both Europe and the USA, and its clinical application in oncology has already been demonstrated (Drzezga et al., 2011). The requirement for MR data has the key benefit that regional volumes and volume changes are also available for each patient, and these data could potentially be combined with the FDG-PET information.

There are two methodological image processing issues which are important to discuss. The first concerns our decision to nonrigidly propagate the baseline segmentations to follow-up space, rather than, for example, using the MAPER segmentation procedure to automatically generate independent follow-up segmentations. Despite the fact that erroneously labelled voxels in the baseline segmentation are propagated to the follow-up image, intra-subject consistency of the segmentation is important for measuring longitudinal change (Crum et al., 2001), since uncorrelated errors lead to greater measurement uncertainty. The second is the issue of FDG-PET image normalisation. The reference cluster normalisation method (Yakushev et al., 2009) was proposed as a data-driven method, with the cluster derived directly from the image data. However, we used an independently derived cluster for normalisation to avoid introducing bias into the classification process. It was important to first assess the validity of this approach, by determining whether the regions identified as relatively preserved in Yakushev et al. (2009) are also valid for the ADNI dataset. We therefore derived a reference cluster using the ADNI FDG-PET images, and calculated the intraclass correlation coefficient (ICC) between the values obtained by sampling this ADNI derived cluster and those obtained by sampling the independently derived cluster. The resulting ICC of 0.95 suggests that the area of the brain identified is reliably preserved across early AD and MCI, and thus is likely to provide a robust and portable reference region for image normalisation.

This study demonstrates that information extracted from serial FDG-PET through regional analysis can accurately discriminate diagnostic groups, even at the early symptomatic stages of the disease. This finding may be usefully applied in the diagnosis of Alzheimer's disease, predicting disease course in individuals with mild cognitive impairment, and in the selection of participants for clinical trials. Importantly, we demonstrate the utility of serial regional FDG-PET for patient classification in a realistic multi-centre setting. Although the use of longitudinal data for the clinical diagnosis of AD is not necessarily practical, its use for stratification of pMCI versus sMCI patients could still be valuable. For clinical trial recruitment, in particular, it may well be acceptable to use longitudinal information acquired over 12 months to gain additional certainty about whether a candidate fits the selection criteria.

We have identified several areas for further research. We have already begun to explore some of the possibilities, such as using a

more sophisticated method for data combination, and making use of both MRI and PET in combination (Gray et al., 2011a). In the future, we intend to additionally investigate the incorporation of non-imaging data, such as CSF biomarkers or genetic information. Machine learning techniques using cross-sectional FDG-PET data have been successful in discriminating AD patients from those with frontotemporal dementia (for example, Kippenhan et al., 1994; Xia et al., 2008), and we would be interested to investigate the effect of incorporating longitudinal information on such differential diagnoses. While it is possible that the ADNI dataset contains some patients with other dementias, such as frontotemporal dementia or dementia with Lewy bodies, these patients are not clinically labelled as such. To perform a thorough study on differential diagnosis, a large and varied cohort of dementia patients with autopsy-confirmed clinical diagnoses would be required, such as that described in Silverman et al. (2001).

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.12.071.

Acknowledgments

This project is partially funded under the 7th Framework Programme by the European Commission (http://cordis.europa.eu/ist).

The authors would like to thank Igor Yakushev (University of Mainz, Germany) for provision of the MNI-space reference cluster image, as well as Casper Nielsen and Kelvin Leung (Dementia Research Centre, University College London) for provision of the MIDAS brain masks. 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 organization 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.

References

ADNI MRI Core, 2005. ADNI MRI technical procedures manual. Available at: http://www.adni-info.org/images/stories//mritrainingmanualv1.pdf 2005.

ADNI PET Core, 2005. ADNI PET technical procedures manual. Available at: http://www.adni-info.org/images/stories/Manuals/pettechnologistmanual-v9.2.pdf 2005.

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.

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.

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

Borghammer, P., Aanerud, J., Gjedde, A., 2009. Data-driven intensity normalization of PET group comparison studies is superior to global mean normalization. NeuroImage 46, 981–988.

Brookmeyer, R., Johnson, E., Ziegler-Grahamm, K., Arrighi, H.M., 2007. Forecasting the global prevalence and burden of Alzheimer's disease. Alzheimers Dement. 3, S168.

Chang, C., Lin, C., 2011. LIBSVM: a library for support vector machines. ACM TIST 2, 27:1–27:27.

Chen, K., Langbaum, J.B.S., Fleisher, A.S., Ayutyanont, N., Reschke, C., Lee, W., Liu, X., Bandy, D., Alexander, G.E., Thompson, P.M., Foster, N.L., Harvey, D.J., de Leon, M.J., Koeppe, R.A., Jagust, W.J., Weiner, M.W., Reiman, E.M., the Alzheimer's Disease Neuroimaging Initiative, 2010. Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative. NeuroImage 51, 654–664.

Chupin, M., Hammers, A., Liu, R., Colliot, O., Burdett, J., Bardinet, E., Duncan, J., Garnero, L., Lemieux, L., 2009. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. NeuroImage 46, 749–761.

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, 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.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Crum, W.R., Scahill, R.I., Fox, N.C., 2001. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. NeuroImage 13, 847–855.

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 6, 766–781.

de Leon, M.J., Convit, A., Wolf, O.T., Tarshish, C.Y., De Santi, S., Rusinek, H., Tsui, W., Kandil, E., Scherer, A.J., Roche, A., Imossi, A., Thorn, E., Bobinski, M., Caraos, C., Lesbre, P., Schlyer, D., Poirieri, J., Reisberg, B., Fowler, J., 2001. Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]flouro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). Proc. Natl. Acad. Sci. 98, 10966–10971.

Drzezga, A., Souvatzoglou1, M., Beer, A., Ziegler, S., Fürst, S., Nekolla, S., Schwaiger, M., 2011. Integrated simultaneous whole-body MR/PET: first comparison between MR/PET and PET/CT in patients. J. Nucl. Med. 52, 262.

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.

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., Aljabar, P., Heckemann, R., Hammers, A., Rueckert, D., 2011a. Random forest-based manifold learning for classification of imaging data in dementia. Machine Learning in Medical Imaging. Springer Berlin/Heidelberg. Volume 7009 of Lecture Notes in Computer Science, pp. 159–166.

Gray, K., Wolz, R., Keihaninejad, S., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., 2011b. Regional analysis of FDG-PET for the classification of Alzheimer's disease. Medical Imaging Understanding and Analysis (MIUA).

Gray, K., Wolz, R., Keihaninejad, S., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., 2011c. Regional analysis of FDG-PET for use in the classification of Alzheimer's disease. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI) 2011, IEEE, pp. 1082–1085.

Haense, C., Herholz, K., Jagust, W.J., Heiss, W.D., 2009. Performance of FDG PET for detection of Alzheimer's disease in two independent multicentre samples (NEST-DD and ADNI). Dement. Geriatr. Cogn. Disord. 28, 259–266.

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.

Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., 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.

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.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G.,

Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., the ADNI Study, 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Jagust, W.J., Bandy, D., Chen, K., Foster, N.L., Landau, S.M., Mathis, C.A., Price, J.C., Reiman, E.M., Skovronsky, D., Koeppe, R.A., Investigators, A.D.N.I., 2010. The ADNI PET core. Alzheimers Dement. 6, 221–229.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 30, 436–443.

Kippenhan, J.S., Barker, W.W., Nagel, J., Grady, C., Duara, R., 1994. Neural-network classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras. J. Nucl. Med. 35, 7–15.

Langbaum, J.B., 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., 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.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). NeuroImage 2, 89–101.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., J.,C.R. Jr., 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.

Minoshima, S., Foster, N.L., Sima, A.A.F., Frey, K.A., Albin, R.L., Kuhl, D.E., 2001. Alzheimer's disease versus dementia with Lewy bodies: cerebral metabolic distinction with autopsy confirmation. Ann. Neurol. 50, 358–365.

Mosconi, L., Perani, D., Sorbi, S., Herholz, K., Nacmias, B., Holthoff, V., Salmon, E., Baron, J.C., De Cristofaro, M.T.R., Padovani, A., Borroni, B., Franceschi, M., Bracco, L., Pupi, A., 2004. MCI conversion to dementia and the APOE genotype — a prediction study with FDG-PET. Neurology 63, 2332–2340.

Mosconi, L., Tsui, W.H., De Santi, S., Li, J., Rusinek, H., Convit, A., Li, Y., Boppana, M., de Leon, M.J., 2005. Reduced hippocampal metabolism in MCI and AD — automated FDG-PET image analysis. Neurology 64, 1860–1867.

Mosconi, L., Tsui, W.H., Herholz, K., Pupi, A., Drzezga, A., Lucignani, G., Reiman, E.M., Holthoff, V., Kalbe, E., Sorbi, S., Diehl-Schmid, J., Perneczky, R., Clerici, F., Caselli, R., Beuthien-Baurnann, B., Kurz, A., Minoshima, S., de Leon, M.J., 2008. Multicenter standardized F-18-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. J. Nucl. Med. 49, 390–398.

Nestor, P.J., Fryer, T.D., Smielewski, P., Hodges, J.R., 2003. Limbic hypometabolism in Alzheimer's disease and mild cognitive impairment. Ann. Neurol. 54, 343–351.

Petersen, R., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 760.

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

Ranginwala, N.A., Hynan, L.S., Weiner, M.F., White III, C.L., 2008. Clinical criteria for the diagnosis of Alzheimer disease: still good after all these years. Am. J. Geriatr. Psychiatry 16, 384–388.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging 18, 712–721.

Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., Illán, I.A., López, M., Segovia, F., Chaves, R., Padilla, P., Puntonet, C.G., 2010. Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. Med. Phys. 37, 6084–6095.

Schnabel, J.A., Rueckert, D., Quist, M., Blackall, J.M., Castellano Smith, A.D., Hartkens, T., Penney, G.P., H.,W.A., Liu, H., Truwit, C.L., Gerritsen, F.A., Hill, D.L.G., Hawkes, D.J., 2001. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 573–581.

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.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, 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.

Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. 32, 71–86.

Thiele, F., Wenzel, F., Young, S., Cross, D., Buchert, R., Minoshima, S., 2009. Metabolic heterogeneity in ADNI PET subjects. J Nucl Med Meetings Abstracts 50, p. 429.

Wolz, R., Heckemann, R.A., Aljabar, P., Hajnal, J.V., Hammers, A., Jyrki Lötjönen, D.R., the Alzheimer's Disease Neuroimaging Initiative, 2010. Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. NeuroImage 52, 109–118.

Xia, Y., Wen, L., Eberl, S., Fulham, M., Feng, D., 2008. Genetic algorithm-based PCA eigenvector selection and weighting for automated identification of dementia using FDG-PET imaging. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS) 2008, IEEE, pp. 4812–4815.

Yakushev, I., Landvogt, C., Buchholz, H.G., Fellgiebel, A., Hammers, A., Scheurich, A., Schmidtmann, I., Gerhard, A., Schreckenberger, M., Bartenstein, P., 2008. Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. Psychiatry Res. 164, 143–153.

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.