Binary classification of ^{18}F-flutemetamol PET using machine learning: comparison with visual reads and structural MRI

Rik Vandenberghe[1,2], Natalie Nelissen[1], Eric Salmon[3], Adrian Ivanoiu[4], Steen Hasselbalch[5], Allan Andersen[6], Alex Korner[7], Lennart Minthon[8], David J. Brooks[9,10], Koen Van Laere[11], Patrick Dupont[1]

[1]Laboratory for Cognitive Neurology, Department of neurosciences, KU Leuven; [2]Neurology Department, University Hospitals Leuven, Belgium; [3]Cyclotron Research Center, University of Liege, Belgium; [4]Neurology Department, Catholic University of Louvain, Brussels, Belgium; [5]Memory Disorders Research Group, Department of Neurology, Rigshospitalet Copenhagen; [6]Glostrup University Hospital, Glostrup, Denmark; [7]Hillerod Hospital, Hillerod, Denmark; [8]Neuropsychiatric Clinic, University Hospital MAS, Lund University, Lund Sweden; [9]GE Healthcare, London, UK; [10]Imperial College, London, UK; [11]Nuclear Medicine Department, University Hospitals Leuven, Belgium

Correspondence to: Rik Vandenberghe, M.D., Ph.D., Neurology Department, University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium, e-mail: rik.vandenberghe@uz.kuleuven.ac.be, phone ++ 32 (0)16 344280, fax: ++ 32 (0)16 3444285

Keywords: amyloid imaging, Alzheimer's disease, mild cognitive impairment, MRI volumetry

NeuroImage (2012), in press
DOI: http://dx.doi.org/10.1016/j.neuroimage.2012.09.015

Abstract

^{18}F-flutemetamol is a positron emission tomography (PET) tracer for in vivo amyloid imaging. The ability to classify amyloid scans in a binary manner as 'normal' versus 'Alzheimer-like', is of high clinical relevance. We evaluated whether a supervised machine learning technique, support vector machines (SVM), can replicate the assignments made by visual readers blind to the clinical diagnosis, which image components have highest diagnostic value according to SVM and how ^{18}F-flutemetamol-based classification using SVM relates to structural MRI-based classification using SVM within the same subjects. By means of SVM with a linear kernel, we analyzed ^{18}F-flutemetamol scans and volumetric MRI scans from 72 cases from the ^{18}F-flutemetamol phase 2 study (27 clinically probable Alzheimer's disease (AD), 20 amnestic mild cognitive impairment (MCI), 25 controls). In a leave-one-out approach, we trained the ^{18}F-flutemetamol based classifier by means of the visual reads and tested whether the classifier was able to reproduce the assignment based on visual reads and which voxels had the highest feature weights. The ^{18}F-flutemetamol based classifier was able to replicate the assignments obtained by visual reads with 100% accuracy. The voxels with highest feature weights were in the striatum, precuneus, cingulate and middle frontal gyrus. Second, to determine concordance between the gray matter volume- and the ^{18}F-flutemetamol-based classification, we trained the classifier with the clinical diagnosis as gold standard. Overall sensitivity of the ^{18}F-flutemetamol- and the gray matter volume-based classifiers were identical (85.2%), albeit with discordant classification in three cases. Specificity of the ^{18}F-flutemetamol based classifier was 92% compared to 68% for MRI. In the MCI group, the ^{18}F-flutemetamol based classifier distinguished more reliably between converters and non-converters than the gray-matter based classifier. The visual read-based binary classification of ^{18}F-flutemetamol scans can be replicated using SVM. In this sample the specificity of ^{18}F-flutemetamol based SVM for distinguishing AD from controls is higher than that of gray matter volume-based SVM.

1 Introduction

Amyloid imaging has the potential to become a clinically important diagnostic tool in the field of Alzheimer's disease (AD) and related disorders. One of the parameters for clinical use that has recently received much attention is the reliability of binary assignments of scans into an 'AD-like' versus a 'normal' pattern. In the phase 2 study of ^{18}F-flutemetamol, the ^{18}F labelled analogue (Koole et al., 2009; Nelissen et al., 2009; Vandenberghe et al., 2010) of ^{11}C-Pittsburgh Compound B (PIB) (Klunk et al., 2003, 2004; Lockhart et al., 2007; Ikonomovic et al., 2008), 27 patients with clinically probable AD, 25 healthy controls and 20 patients with amnestic mild cognitive impairment (MCI) participated. Five independent readers who were blind to the clinical diagnosis assigned each of the 72 scans to a 'raised' or a 'normal retention' category. In case of disagreement the decision was based on a majority verdict. In all AD cases, readers concurred in their assessment except for one reader in one AD case. Likewise, scans in all healthy volunteers were rated consistently between the 5 readers (Vandenberghe et al., 2010). Among the MCI cases, 9 were assigned to the 'raised retention' and 11 to the 'normal retention' category (Vandenberghe et al., 2010). A first purpose of the current study was to evaluate whether a supervised machine learning technique, support vector machines (SVM) with a linear kernel (Klöppel et al., 2008b), could be trained to replicate the outcome of the visual reads and which image components contributed most to the binary SVM classification.

In SVM with a linear kernel, the algorithm is trained to classify scans into two a priori defined classes (e.g. patients vs controls). During training the algorithm tries to find the hyperplane that optimally separates (maximizes the margin between) the two classes (Fig. 1). Subsequently, performance is tested using a test set. The outcome parameters of SVM are based on the accuracy of classification on the individual case level. Although other classification techniques such as logistic regression, linear discriminant analysis, neural networks, ... can be used, we opted for SVM because it is well established as a classification technique for neuroimaging data (Davatzikos et al., 2008; Klöppel et al., 2008a,b; Vemuri et al., 2008; Magnin et al., 2009; Plant et al., 2010; Chincarini et al., 2011; Padilla et al., 2012). This allowed us to directly compare performance of a same classifier method between the MRI and the PET modality. A key advantage of SVM over other methods is that it allows inferences to be made at the level of the individual (Orrù et al., 2012). Furthermore, it can handle huge sets of values per case in a multivariate way and is therefore ideally suited for analyzing images consisting of 10,000s of voxels representing a given feature without need for volume of interest-based data reduction. From a clinical standpoint, it is important to empirically define to which degree in vivo amyloid imaging and structural MRI within the same subjects provide redundant, complementary or contradictory information. The second aim therefore of the current study was to evaluate the degree of concordance of binary SVM classifications based on either ^{18}F-flutemetamol or gray matter volume maps.

To address our primary question, we used the visual reads to define the 2 a priori classes and train the algorithm. We compared the classification of the test cases with that based on visual reads. We will also present a voxel-based map of topographical distribution of the feature weights. Feature weights indicate how much each voxel contributes to the pattern that is used by the classifier to separate the two groups. To address our second question, we used the clinical diagnosis for definition of the a priori classes in order not to bias the algorithm in favor of one or the other imaging technique. We evaluated how concordant the classification was between structural MRI and ^{18}F-flutemetamol.

2 Subjects and Methods

2.1 Subjects

Our dataset consisted of the 72 ^{18}F-flutemetamol and structural MRI scans from the ALZ201 phase 2 study (Vandenberghe et al., 2010). In this study, 27 patients who fulfilled the National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association criteria for clinically probable AD and the Diagnostic and Statistical Manual of Mental Disorders-IV criteria for dementia of the Alzheimer type (mean age: 69.6 years, S.D. 7.03; Mini Mental State Examination (MMSE) score 23.3, S.D. 2.18) participated, 20 patients with amnestic MCI (Petersen, 2004) (mean age: 72.7 years, S.D. 7.09, MMSE 28.0, S.D. 0.94), 15 cognitively intact controls above the age of 55 (mean age: 68.7 years, S.D. 7.61; MMSE 29.4, S.D. 0.97) and 10 cognitively intact controls below the age of 55 (mean age: 37.9 years, S.D. 11.5; MMSE 29.4, S.D. 0.92). Inclusion and exclusion criteria and demographic variables of this study sample has been described previously (Vandenberghe et al., 2010). Patients were recruited through academic memory clinics and controls through advertisement in local newspapers.

2.2 Image acquisition

2.2.1 ^{18}F-flutemetamol PET

At two radiopharmaceutical production sites, in Belgium and in Denmark, the investigational medicinal product ^{18}F-flutemetamol was batch manufactured according to good manufacturing practice guidelines (EudraLex, volume 4, annex 3) using a TracerLab-FXF-N chemistry platform (GE Healthcare). Subjects received a dynamic ^{18}F-flutemetamol PET scan consisting of six five-minute frames recorded during the window at 85-115 minutes post injection. Dynamic brain scanning was performed at 3 different scanning centers using a 16-slice Biograph PET/CT scanner (Siemens, Erlangen, Germany), an ECAT EXACT HR+ scanner (Siemens, Erlangen, Germany), and a GE Advance scanner. The reconstruction method consisted of filtered back projection in case of the HR+ and Advance scanners and iterative reconstruction (ordered-subset expectation maximization) in the case of the Biograph 16. Scans were reconstructed as 6 5-min frames (for details see (Vandenberghe et al., 2010)). Further details are described in the original phase 2 paper (Vandenberghe et al., 2010).

2.2.2 Volumetric MRI

T1-weighted magnetic resonance images were acquired using a standard three-dimensional Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) volumetric MRI scan at 3T (1 mm isotropic voxels). Scans were acquired at 4 different MR scanning centres, two of which used a Philips Achieva MRI (12 AD, 15 MCI, 20 controls), with the two remaining centres using a Siemens Allegra (7 AD), and Siemens Magnetom Trio (8 AD, 5 MCI, 5 controls), respectively. The maximum interval between the PET and the MRI was 30 days. Subjects also received a Fluid Attenuated Inversion Recovery (FLAIR) MRI sequence.

2.3 Image preprocessing

2.3.1 ^{18}F-flutemetamol PET

All PET frames were realigned, and a PET summed image was created. We calculated standard uptake value ratio (SUVR) images from the PET summed image using manually delineated cerebellar gray matter as reference region (Nelissen et al., 2009; Vandenberghe et al., 2010). The PET summed image was coregistered to the subject's T1-weighted MRI using a mutual information-based method (Maes et al., 1997). The reoriented MRI was spatially normalized into Montreal Neurologic Institute (MNI) space. The resulting transformation was then applied to the coregistered PET summed images. The voxel size of the normalized SUVR images was 2x2x2 mm^3.

2.3.2 Volumetric MRI

MRI images were segmented into gray matter and spatially normalised using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/). In this procedure the warping to MNI is performed using Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL) (Ashburner, 2007) and a bias field correction is applied prior to the segmentation. Gray matter images were modulated in order to create gray matter volume maps. All images were masked using the SPM8 brain mask excluding the ventricles. The voxel size of the final images was 1.5 x 1.5 x 1.5 mm^3.

2.4 Image analysis

We used the machine learning package The Spider version 1.71 (http://people.kyb.tuebingen.mpg.de/spider/, Weston J., Elisseef A, BakIr G., and Sinz F., Max Planck Institute for Biological Cybernetics, Tübingen, Germany), running under Matlab version 7.9 to calculate the support vector machine (SVM) with a linear kernel using the default soft margin option. Each image represents a single point in a high dimensional space in which each dimension corresponds to one voxel within the brain mask. The coordinate is the intensity value in that voxel. The weight of a feature (voxel) is the component of the normal vector of the hyperplane along that voxel's dimension (Fig. 1). It indicates how much that voxel contributes to the pattern underlying overall classification. The feature weights of the classifier will be presented as an image in MNI space.

2.4.1 Comparison with visual reads

In a leave-one-out approach we used the scans from the 25 AD subjects which visual readers had previously assigned to the 'raised retention' category ('positive' read) and scans from the 24 controls which visual readers had previously assigned to the 'normal retention' category ('negative' read) (Vandenberghe et al., 2010). The SVM was trained to classify a scan into one of two classes, 'AD-like' versus 'normal retention' pattern, using all scans less one. The test set consisted of the remaining scan. By means of permutations we cycled through the entire set so that each scan was used once as the test set. In addition to the simple binary classification into 'raised' or 'normal retention', we will also report the distance of each scan to the separating hyperplane: Larger absolute distances indicate a more robust classification.

In a second approach, the training set consisted of the full set of positively read scans from 25 AD subjects and the negatively read scans from 24 controls. The independent

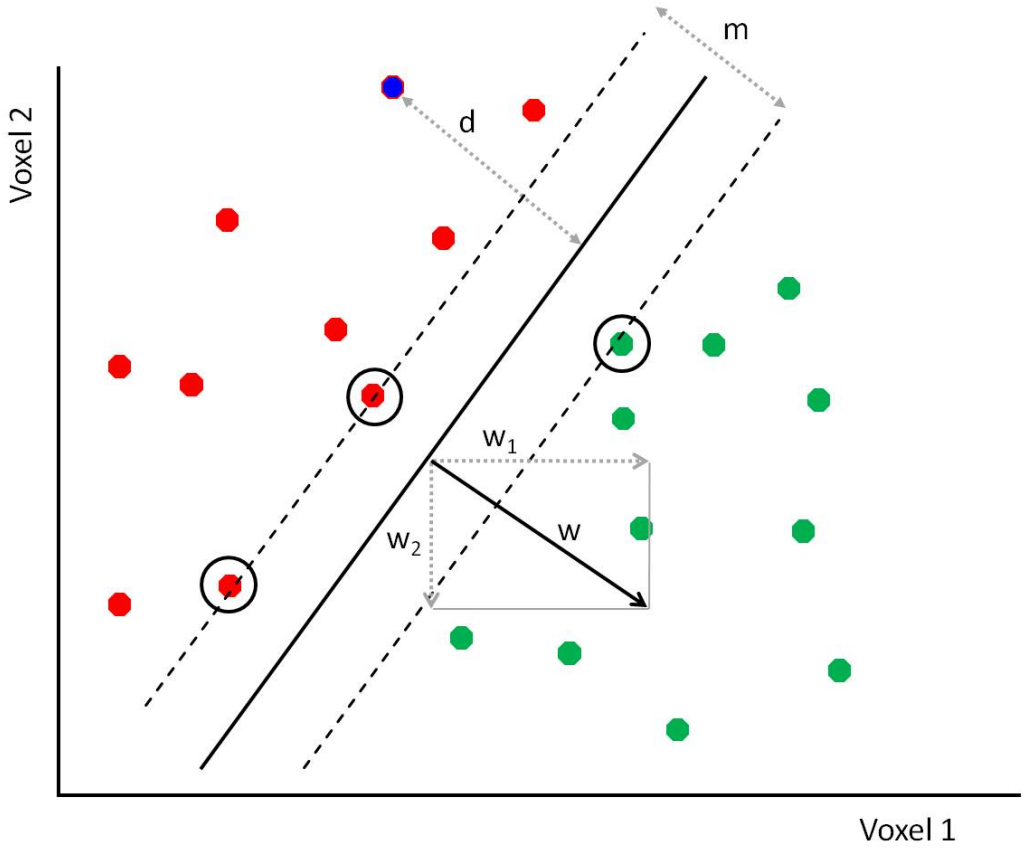

Figure 1: Illustration of a linear support vector machine using 2 features, in this instance voxels 1 and 2. In the current whole-brain PET study, each voxel can be thought of as 1 dimension in a multidimensional space. *Legend*: red: training cases belonging to class A; green: training cases belonging to class B; blue: test case. Circled dots correspond to the support vectors; m = margin, i.e. distance between support vectors; d: distance from a case to the hyperplane; w = normal vector of the hyperplane, components w_1 and w_2 are the weights of voxels 1 and 2.

test set consisted of all remaining ^{18}F-flutemetamol scans from the phase 2 study: scans from 20 MCI subjects, from two AD subjects which readers had assigned to the normal retention category and from one control which readers had assigned to the raised retention category. In order to illustrate which brain components have most discriminative value based according to this classifier, the feature weights for each voxel will be shown on a normalized brain MRI.

2.4.2 Comparison with gray matter volume-based classifier

To compare an ^{18}F-flutemetamol-based with a gray matter volume-based classifier, we trained the classifier on the basis of the clinical categories: clinically probable AD (n = 27) versus controls (n = 25), for each modality separately. First, we applied a leave-one-out approach. We compared diagnostic parameters (sensitivity and specificity) for an ^{18}F-flutemetamol-based SVM versus a gray matter volume-based SVM classification. Next, we evaluated concordance on a case-by-case level. To assess the statistical significance of the distance to the hyperplane in each case for each modality, a bootstrap approach was used based on 1000 random assignments of the scans of the training set to one of the two classes (AD-like versus normal). From the probability distributions we determined the statistical significance of the distance in each test case, applying a threshold of $P < 0.05$.

Second, to compare performance in the MCI cases, using the 27 clinically probable AD cases and the 25 controls as training set, we applied the classifier to the set of 20 MCI subjects. We compared performance between the ^{18}F-flutemetamol-based and the gray matter volume-based classifier. We also evaluated concordance on a case-by-case level. We also compared the outcome with that of the visual reads of the ^{18}F-flutemetamol scans (Vandenberghe et al., 2010).

Per protocol a 2-year study visit was to be performed in all participants. The study physician had to report whether the diagnosis had changed. We will also report the outcome of this 2-year follow-up visit in the MCI cases. All MCI cases were re-evaluated at this follow-up visit except one.

2.4.3 Comparison with SPM analysis of between-group differences

We also evaluated to which degree the topographic pattern of feature weights corresponded to regional differences between the groups in amyloid burden or gray matter volume. We performed an SPM analysis between the AD group and the normal control group using the PET amyloid images or using the GM volume images. As a standard preprocessing step, we applied an isotropic Gaussian smoothing of the images using a kernel of $FWHM = 8mm$. As statistical threshold we used an uncorrected $P < 0.001$ at the voxel level combined with a FWE (family wise error) corrected $P < 0.05$ at the cluster level.

3 Results

3.1 Comparison with visual reads

The leave-one-out approach fully replicated the classification based on the visual reads (sensitivity and specificity of 100%) (Fig. 2A). Likewise, when we trained the algorithm with the 25 scans from AD subjects and the 24 scans from controls and applied the algorithm to an independent test set consisting of the remaining 23 scans from the ALZ201 study, the classifier reached a specificity and a sensitivity of 100% compared to the visual reads as gold standard (Fig. 2B).

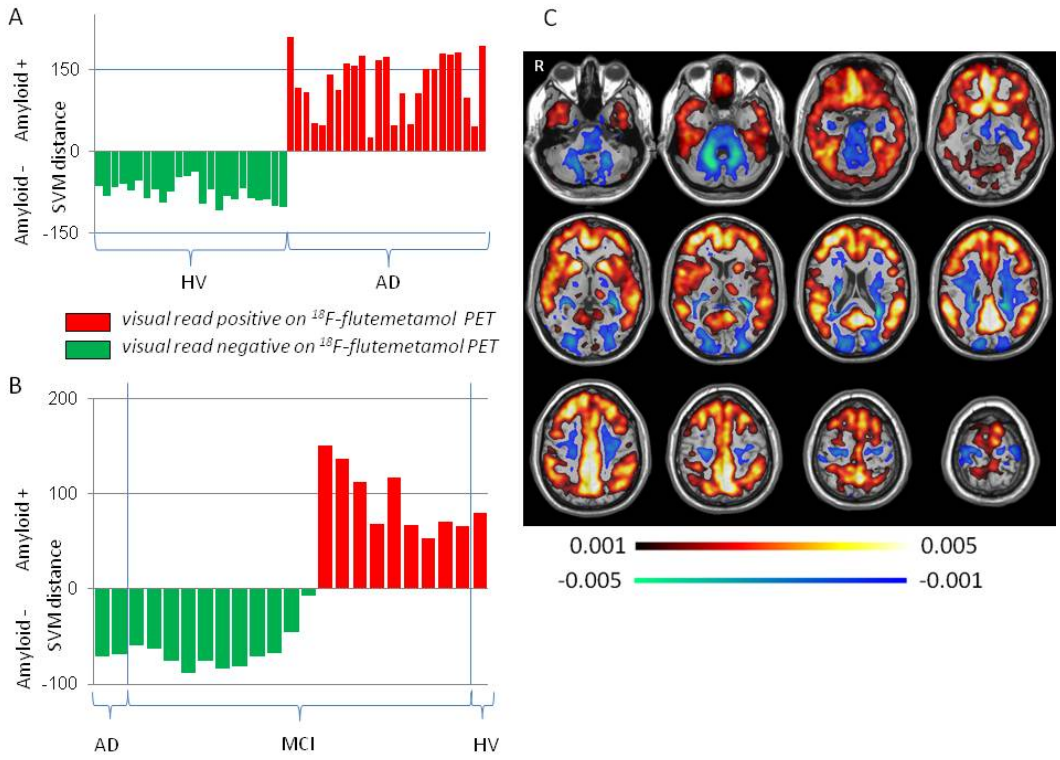

Figure 2: ^{18}F-flutemetamol-based classification of ALZ201 cases. A. Plots of distance of the test subjects to the hyperplane using a leave-one-out approach based on 24 healthy controls with negative reads (green color) and 25 probable AD cases with positive reads (red color). B. Independent test set with ^{18}F-flutemetamol scans from two probable AD cases with negative read, 20 MCI cases and one healthy control with positive read. *Legend:* X axis: individual cases. Y axis: distance from the hyperplane (arbitrary units). The sign of the y value, positive or negative, corresponds to the class to which the case is assigned by the algorithm. Negative corresponds to 'normal' retention, positive to 'AD-like' but the sign of the Y value is otherwise arbitrary. Green: read as 'normal retention' by the readers blind to the diagnosis. Red: read as 'raised retention' by the readers. C. Feature weights of the classifier projected onto a normalized structural MRI in MNI space. Higher positive weights indicate that higher SUVR values in this voxel contribute more toward a classification as 'raised retention', whereas negative weights indicate that higher SUVR values in that voxel results in a higher likelihood of classification as 'normal retention'. Scaled between -0.005 and -0.001 (blue color range) and +0.001 and +0.005 (hot color range). The scale of the feature weights is representing how much a voxel is contributing. We have normalized this scale such that the sum of all weights is one. Only voxels with a weight more than 0.001 (in absolute value) are shown.

Positive feature weights contributing to a 'raised retention' classification (Fig. 2C) were highest in precuneus, striatum, posterior and mid cingulate, anteromedial frontal cortex and middle frontal gyrus. Negative feature weights contributing to a 'normal retention' classification were highest in cerebellum and periventricular white matter.

A classifier is determined by all voxels and their corresponding weights. To show that voxels with a low weight can be neglected in the classification process without loss of performance, we constructed a new classifier using only those voxels with an absolute weight higher than 0.005 (2990 voxels). This new classifier reached an accuracy that was as high as when the original classifier was used.

When we excluded the scans from the 10 controls below the age of 55, specificity and sensitivity remained the same for both approaches (100%).

3.2 Comparison with gray matter volume-based SVM in AD and healthy controls

Using a leave-one-out approach with the clinical diagnosis as gold standard in the 27 AD subjects and the 25 controls, specificity was 92% and sensitivity 85.2% for the ^{18}F-flutemetamol scans (Fig 3A). The distance to the hyperplane was significant in nearly all cases (Fig 3A). For the gray matter volume maps, specificity was 68% and sensitivity 85.2% (Fig 3B). The distance to the hyperplane was significant in most of the normal controls. It failed to reach significance in any of the AD cases in the current sample according to the bootstrapping method (Fig 3B).

Among the scans from clinically probable AD cases, the ^{18}F-flutemetamol-based SVM classified 4 cases as normal, two of whom with a relatively large distance from the hyperplane. The latter two corresponded to those that had also been assigned by the visual readers to the 'normal retention' category. The ^{18}F-flutemetamol-based SVM classified two healthy controls as 'AD-like', one with a small and the other with a larger distance from the hyperplane (Fig 3A). The latter case had also received a positive read of the ^{18}F-flutemetamol scan (Vandenberghe et al., 2010).

Among the scans from healthy controls, the gray matter volume-based SVM classified 8 cases as 'AD-like', in two of whom there was a very small distance from the hyperplane (Fig 3B). The gray matter volume-based classifier also assigned 4 clinically probable AD cases to the normal category, two of whom corresponded to the two AD cases classified as 'normal' based on the ^{18}F-flutemetamol-based SVM.

The highest feature weights for the gray matter volume-based classifier were in medial temporal cortex bilaterally, posterior temporal cortex bilaterally, and medial parietal cortex (Fig. 4).

To determine how subject differences in total gray matter volume affected performance of the classifier, we performed an additional analysis where we scaled the GM volume with total GM. Scaling changed the classification of three cases. These cases corresponded to the three cases who had a GM volume image with the smallest distance to the classifier in the original analysis. This scaling led to a reduction in sensitivity to 74.1%. Specificity remained the same.

In a secondary analysis, when we excluded the scans from the 10 controls below the age of 55, specificity of the ^{18}F-flutemetamol based classifier was 86.7% and sensitivity remained 85.2%. Sensitivity of the gray matter volume-based classifier was 92.6% but specificity further reduced to 20.0%.

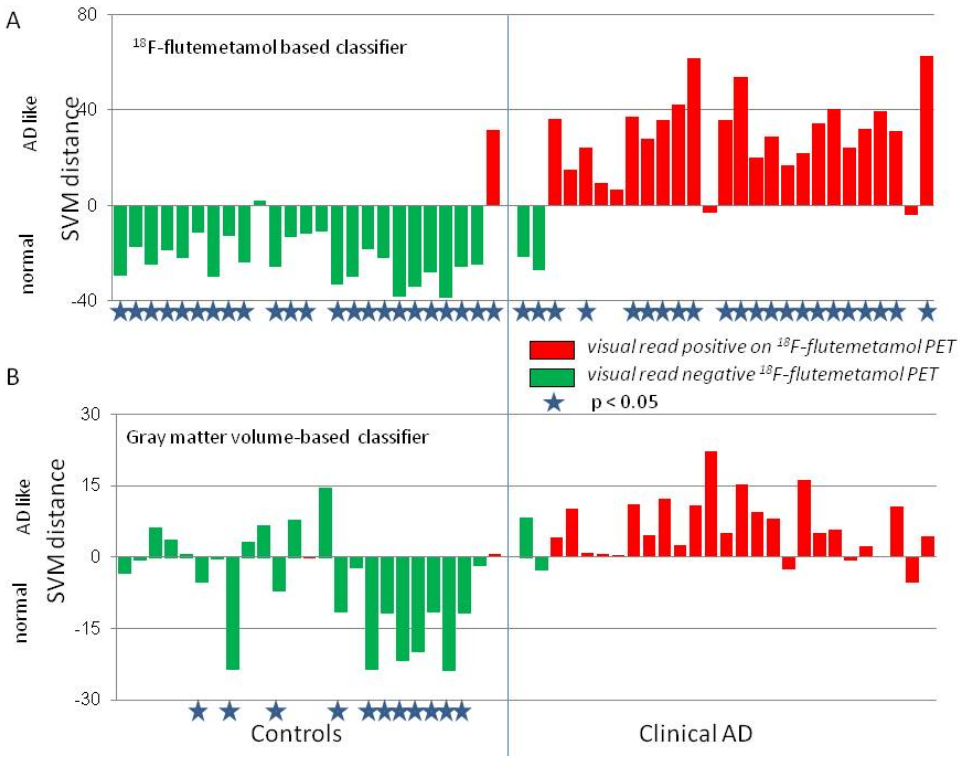

Figure 3: Comparison between ^{18}F-flutemetamol- and gray matter volume-based classifier in 27 clinically probable AD and 25 controls. Leave-one-out approach with clinical diagnosis as gold standard. Significance was determined using a bootstrap method. A star indicates which cases were significant ($P < 0.05$). A. Distance to the hyperplane of the classifier based on ^{18}F-flutemetamol scans (y axis). B. Distance to the hyperplane of the classifier based on gray matter volume maps (y axis) as input.

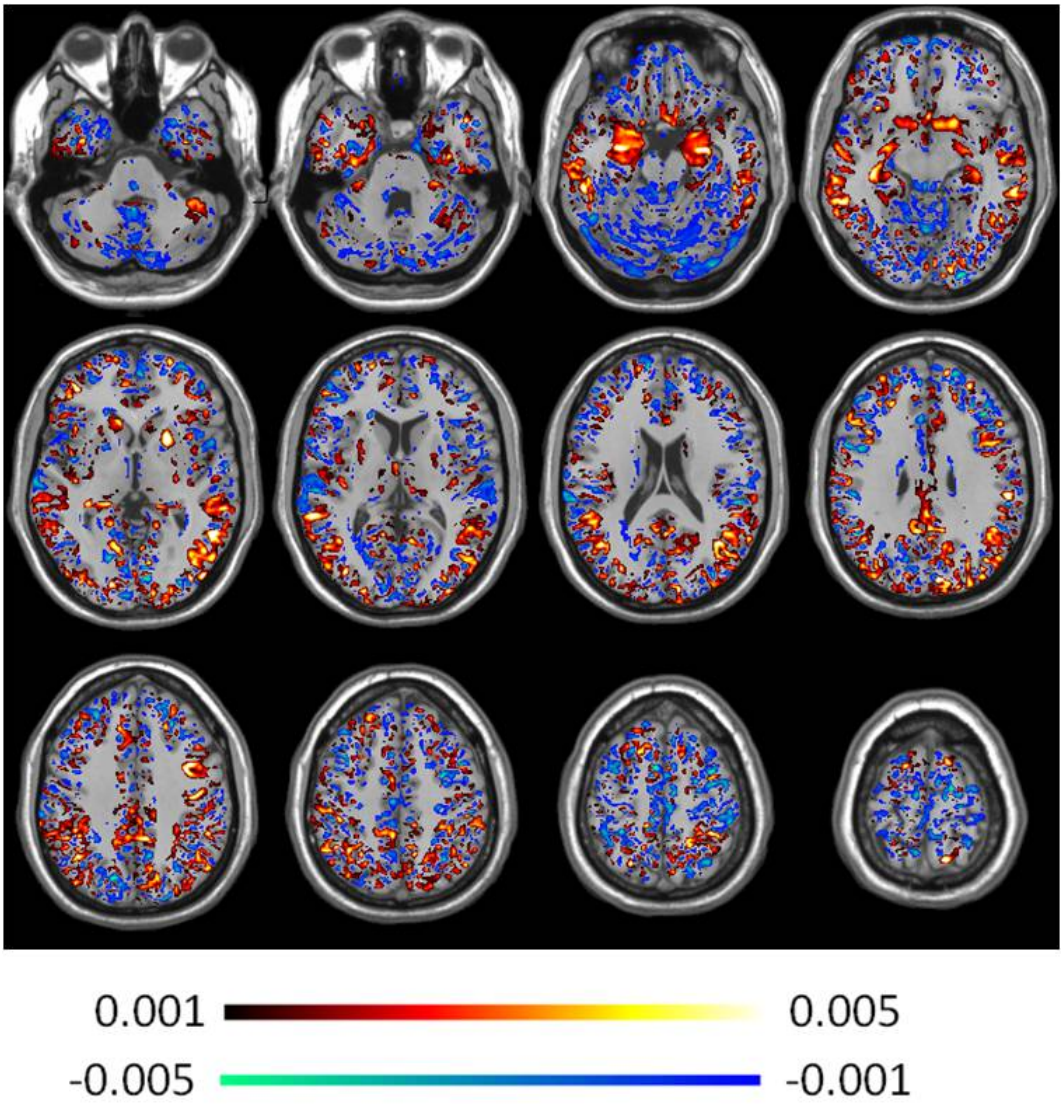

Figure 4: Feature weights of the classifier projected onto a normalized structural MRI in MNI space. Classifier weights for the gray matter volume-based SVM. Voxels with a high feature weight in favor of an 'AD like' classification are shown in hot colors, voxels with a high feature weight in favor of a normal classification are shown in cold colors. Scaled between -0.005 and -0.001 and +0.001 and +0.005. We have normalized this scale such that the sum of all weights is one. Only voxels with a weight more than 0.001 (in absolute value) are shown.

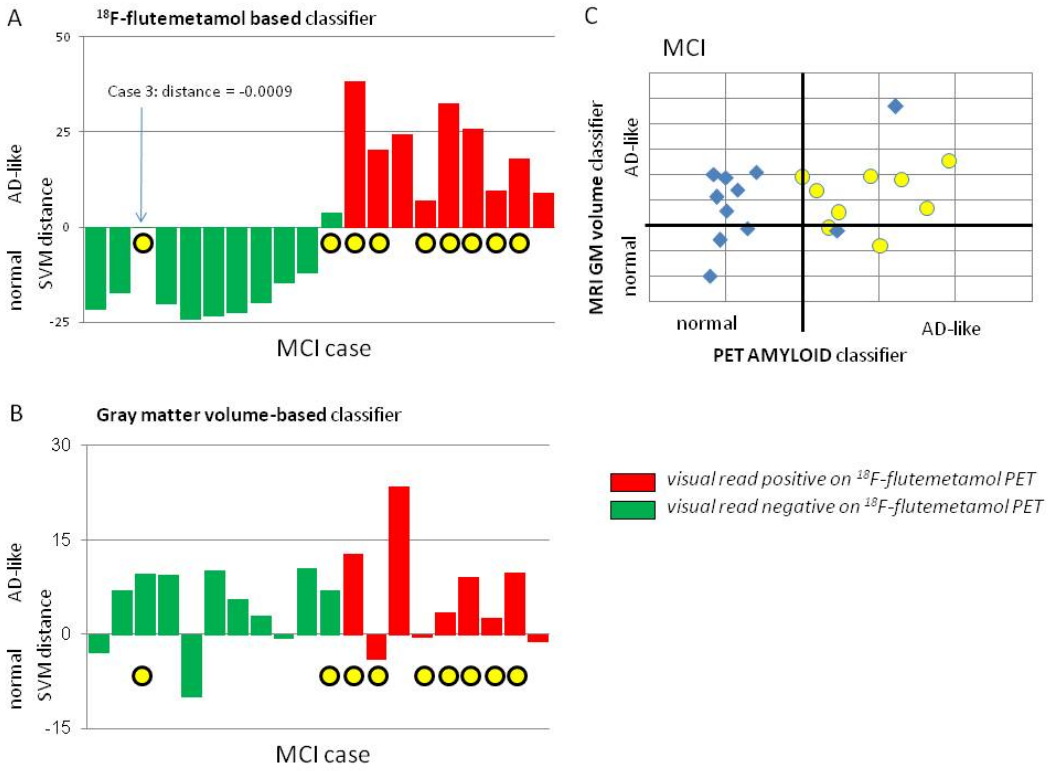

Figure 5: MCI cases (n = 20). The 9 cases who converted to clinically probable AD at 2 years follow-up are marked with a yellow circle. **A.** X-axis: MCI cases. Distance to the hyperplane of the classifier determined with ^{18}F-flutemetamol scans (y axis). *Legend:* Green: cases in whom the ^{18}F-flutemetamol scan was read as negative, red: cases in whom the ^{18}F-flutemetamol scan was read as positive. **B.** Distance to the hyperplane of the classifier determined with gray matter volume maps (y axis) as input. **C.** Distance to the hyperplane of the classifier determined with ^{18}F-flutemetamol scans as input (x axis) and with gray matter volume maps as input (y axis). Filled yellow circles: cases who converted to clinically probable AD within the first 2 years following the PET according to the clinician's judgement.

3.3 Comparison with gray matter volume-based SVM in MCI

According to the clinician's judgement at the 2-year follow-up visit, nine of the 20 MCI cases had converted to clinically probable AD. Of the 10 cases which the ^{18}F-flutemetamol-based classifier categorized as 'AD-like', 8 converted to AD at the prespecified 2-year follow up evaluation (Fig. 5A, yellow circles). One of these cases had initially been read as 'normal retention' by the visual readers (Fig. 5A, green bar with yellow circle). Among the 14 cases which the gray matter volume-based classifier categorized as 'AD-like', 7 converted to AD within 2 years (Fig. 5B, yellow circles). Two additional cases converted which the gray matter volume-based classifier had categorized as 'normal' (Fig. 5B,C).

One MCI case who received a clinically probable AD diagnosis at 2 years follow-up was situated very close to the hyperplane of the ^{18}F-flutemetamol-based classifier (Fig. 5A, 3rd column) but was classified as AD-like according to the gray matter volume-based classifier, with a relatively large distance from the hyperplane (Fig. 5B, 3rd column). This case also received a cerebrospinal fluid biomarker analysis around the same period as the PET and MRI. Aβ42 was within the normal range (743 pg/ml) while [181]phosphotau was increased above the normal range (122 pg/ml) and total tau was at a borderline

level (296 pg/ml) (Hort et al., 2010). At the time of scanning this subject had an isolated amnestic syndrome that progressed steadily thereafter, with additional language and executive dysfunction abnormalities. After 3 years of follow-up the clinical picture remained predominantly cognitive with appearance of distinct comportmental changes including repetitive behavior (continuous humming, repetition of stereotyped sentences) and mild euphoria. A fluorodeoxyglucose PET 3 years after the study scans revealed pronounced left frontal hypometabolism, according to a pattern diagnostic for frontotemporal lobar degeneration (FTLD). This case currently fulfills the criteria for clinically probable behavioral variant frontotemporal degeneration (Rascovsky et al., 2011).

3.4 Comparison with SPM analysis of between-group differences

Significant differences in amyloid ligand retention between the AD group and the normal control group are shown in Fig. 6 (A,B). A similar analysis of the GM volume maps revealed significant decreases in GM volume in medial temporal cortex, the middle temporal gyrus bilaterally, left angular gyrus, posterior cingulate and precuneus (Fig. 6C).

4 Discussion

An SVM can be trained to make binary assignments of ^{18}F-Flutemetamol scans in probable AD, amnestic MCI and controls that closely match those made by visual readers. This provides supportive evidence for the reliability of binary classification of ^{18}F-flutemetamol scans by visual readers in these populations. Image components that have highest discriminative value are located in precuneus, striatum, bilateral middle frontal gyrus and frontomedial cortex. In a direct within-subjects comparison between ^{18}F-flutemetamol and gray matter volume maps in AD and controls, the specificity of ^{18}F-flutemetamol was substantially higher than that of MRI. Likewise, in amnestic MCI the gray matter volume-based classifier assigned more cases to the AD-like pattern than the ^{18}F-flutemetamol based classifier.

In the current study, we made use of the structural MRI to calculate the parameters for normalizing the PET to MNI space scans but in a clinical context, other templates, e.g. based on ^{18}F-flutemetamol, may become available so that MRI would not be a prerequisite for this approach. We used the visual reads for training the classifier. Binary categorization of cases based on semi-quantitative assessment would have yielded identical results given the strict concordance between visual reads and semi-quantitative assessment in the phase 2 data of AD and controls (Vandenberghe et al., 2010).

In this phase 2 study sample, SVM was able to fully replicate the results from the visual reads when the classifier was trained based on the visual reads. This corroborates the validity of the visual reads but does not allow to determine the added value of SVM compared to visual reads. In the second part, performance of the two approaches was compared to the clinical diagnosis and there was again a high concordance between classification based on visual reads versus SVM of ^{18}F-flutemetamol scans. The only exception was one MCI converter who had been read as negative but was classified by the flutemetamol-based classifier as AD-like (Fig. 5). The introduction of machine-intelligence usually improves on the results from human operators. When however comparing the phase 2 study sample of clinically probable AD patients versus cognitively intact controls, visual reads are performing so well (sensitivity and specificity of 93% (Vandenberghe et al., 2010)) that SVM can only add little value. Performance of visual reads however may be different in other

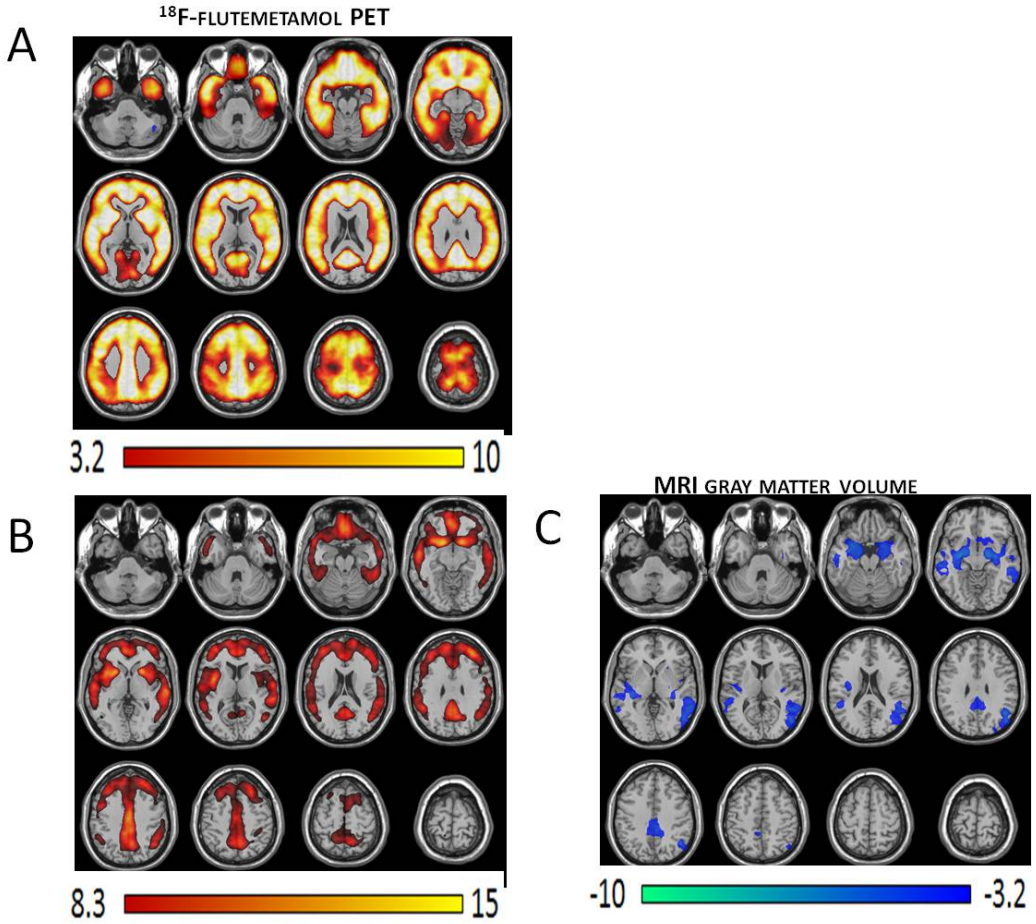

Figure 6: SPM analysis comparing the AD group and the normal control group. SPM maps are thresholded and projected onto a normal MRI in MNI space. A,B. PET amyloid: voxels showing a significant increase in SUVR value in the AD group compared to the normal controls. A. thresholded at uncorrected $P < 0.001$ at the voxel level combined with a FWE corrected $P< 0.05$ at the cluster level. B. thresholded at FWE corrected $P<10^{-6}$ at the voxel level. C. MRI GM volume maps: voxels showing a significant decrease in GM volume in the AD group compared to the normal controls. Thresholded at an uncorrected $P < 0.001$ at the voxel level combined with a FWE corrected $P < 0.05$ at the cluster level.

samples, e.g. subgroups of normal controls with SUVR values close to threshold or pathological cases with atypical presentations such as logopenic aphasia, with values closer to threshold (Leyton et al., 2011) than typical clinically probable AD cases. In samples with higher proportions of cases with near-threshold values, the discordance between visual reads and SVM may rise, and so may the added value of the pattern classifier.

^{18}F-flutemetamol has been introduced relatively recently (Nelissen et al., 2009; Vandenberghe et al., 2010) and the amount of scans that we have available in AD patients and healthy controls is still relatively limited. For that reason, it is impossible to evaluate the diagnostic performance of the ^{18}F-flutemetamol-based classifier in a large test set that is strictly independent of the training set. Furthermore, given the recency of the phase 2 study, we do not have autopsy confirmation in any of the cases included. The two-year conversion rate in our amnestic MCI might serve as a comparator, but in the phase 2 study the evaluation at follow-up was made by clinicians who where not blind to the initial imaging results, but was based upon their clinical judgement informed by the earlier findings, and should thus be interpreted with caution. This is also the main reason why we did not train the classifier based on the 2-year conversion data to avoid circularity. Perhaps the two most relevant cases in this respect are those in whom the clinician diagnosed a conversion to probable AD despite the negative reads because the decision in these individual cases goes against the bias that the unblinded procedure could have induced. It is also worth noting that the gray matter volume-based classifier assigned one of these two cases to the AD-like group, mainly because of substantial hippocampal atrophy. This case (case 3, Fig. 5) evolved into a clinically probable frontotemporal dementia phenotype, with AD-biomarkers on CSF negative for AD. FTLD cases may present with an amnestic syndrome (Hodges et al., 2004; Knopman et al., 2005) and significant hippocampal involvement (Pao et al., 2011), in which case the MRI-based classifier may assign the scan to an 'AD-like' pattern but not the ^{18}F-flutemetamol-based classifier. The second converter (case 11, Fig. 5) who had been read as negative was classified by both the PET- and the gray matter volume-based classifer as AD-like but with a distance very close to the hyperplane (Fig. 5).

Compared to the number of voxels included in each image volume, the number of cases included in the training set is relatively low due to the limited number of ^{18}F-flutemetamol scans available at the time of writing. This could raise the criticism of a mismatch between the sample size and the number of variables (voxels) used for classification. It is, however, important to note that SVM is not a voxel-by-voxel approach and that the classifier defines the hyperplane on the basis of a pattern. The mismatch between sample size and number of discrete voxels within each image is therefore less critical for a pattern classification than for a univariate approach.

The distribution of cases belonging to the different study groups between the different centres in the phase 2 study was such that the current dataset does not lend itself well to state-of-the-art systematic study of between-centre and between-scanner effects, their origin, and their effect on case classification. Given the use of amyloid imaging for clinical multicentre trials, this remains an important topic for further investigation. In one retrospective large European multicentre study, the between-centre variability of ^{11}C-PIB scans was relatively limited (Nordberg et al., 2012).

The main purpose of the current study was to evaluate binary assignment of images based on discrete diagnostic categories. An alternative approach would be a regression analysis to evaluate the independent contribution of continuous variables such as cognitive test scores, age, education level etc. on ligand retention (Pike et al., 2011; Chételat et al., 2010, 2011) and gray matter volume. This complementary approach however is outside the scope of the current paper. We compared the distribution of the feature weights to

the anatomy of the between-group differences as determined by a classical SPM-based group contrast. Overall, regions with high classification weights corresponded relatively well to regions containing significant differences between the two groups (Fig. 6A,B). In a clinical context, the diagnostic classification at the individual level, e.g. by means of a classifier, is of the essence rather than differences based on group comparisons.

The highest feature weights of the ^{18}F-flutemetamol-based classifier were mostly located where one would expect from the previous literature on beta-amyloid tracers, such as precuneus, cingulum and middle frontal gyrus (Nelissen et al., 2009; Vandenberghe et al., 2010) (Fig. 2C). The high feature weights in striatum were not expected. This finding may reflect that the striatum is typically less attended in visual reads of beta-amyloid PET, in part because of imperfect delineation of the gray and white matter which makes the boundary more difficult to evaluate at the level of the subcortical structures than the cortical windings. The striatum is loaded with amyloid plaques in AD (Rudelli et al., 1984; Gearing et al., 1993; Brilliant et al., 1997) and the ventral striatum is particularly vulnerable (Suenaga et al., 1990). In the dorsal striatum the plaques are predominantly of the diffuse type (Gearing et al., 1993) while in the ventral striatum, where we found the highest feature weights, the plaques are associated with dystrophic neurites (Suenaga et al., 1990). The striatal plaques are immunohistochemically distinct from cortical diffuse or neuritic plaques (Gearing et al., 1993) and have been considered a reasonably reliable histopathological marker for AD (Gearing et al., 1993).

The high feature weights in white matter may seem puzzling at first since retention in this area does not differ between the control group and the AD cases (Vandenberghe et al., 2010), and hence might not be expected to influence the classification much. However, the feature weights do not express how well an individual voxel on its own allows one to discriminate between the two groups, but rather are an expression of how much it contributes in a given voxel to the pattern that allows the classifier to discriminate between groups. This means that a voxel in white matter may have a high feature weight if in combination with other voxels it leads to optimal classification of cases. A plausible explanation for the emergence of high feature weights in white matter is that the SUVR PET values of the boundary between white matter and gray matter contain important discriminative information.

The map of feature weights for the MRI-based classifier is consistent with what is found in other SVM studies, with high values in amygdala, hippocampus, lateral temporal cortex and around the posterior cingulate sulcus (Vemuri et al., 2008; Liu et al., 2012). Our gray matter volume-based classifier had a specificity of 68%, compared to 92% for the ^{18}F-flutemetamol-based classifier. A larger proportion of healthy elderly controls was classified as AD-like by MRI. When the hippocampi in this dataset are segmented using an extended multi-atlas segmentation method (Lötjönen et al., 2011) and volumes are quantified, hippocampal volumes overlap between a significant portion of healthy controls and AD cases (Thurfjell et al., 2012). A similar degree of overlap is found when hippocampal volumes are rated visually (Duara et al., 2012). As the hippocampus received high feature weights in our gray matter volume-based classifier (Fig. 4), we can probably partly account for the low specificity of the gray matter volume-based classifier due to the overlap in medial temporal volume between the healthy controls and the AD subjects. Other SVM studies of MRI in AD and controls yielded sensitivity values ranging between 61 and 86% and specificity values ranging between 80 and 95% (Klöppel et al., 2008b; Vemuri et al., 2008; Magnin et al., 2009; Liu et al., 2012). Our SVM approach was identical to that used by Klöppel et al. (2008). If the outcome of classifiers is compared between studies, it is important to ensure the comparability of the groups (e.g. regarding stage and age). Our sample is an early-stage AD sample in subjects below the age of

80. Other factors that could theoretically have negatively affected performance of the MRI classifier in our study is the usage of scans acquired from 4 different centres and 3 different MRI scanners, without strict procedures to ensure between-centre replicability. A recent study (Abdulkadir et al., 2011) reported effects of hardware heterogeneity on the performance of MRI based SVM classification in AD. They found that a change in hardware could lead to a change of the decision value. Therefore, the results are expected to improve when performed in only one center on a single scanner. This said, the ^{18}F-flutemetamol scans were also acquired at multiple centres and specificity for this modality was clearly higher than for MRI. It is possible that other MRI features, such as cortical thickness, would provide a higher degree of concordance between the PET- and the gray matter volume-based classifier (Becker et al., 2011). More advanced techniques for feature selection and extraction (Liu et al., 2012), use of nonlinear kernels (Vemuri et al., 2008) or adding variables such as age and gender into the classifier model (Vemuri et al., 2008) may also improve performance. The purpose however of the current study was to compare a same mathematical approach applied to either ^{18}F-flutemetamol PET or structural MRI within the same subjects.

FDG-PET was not part of the ^{18}F-flutemetamol phase 2 study protocol (Vandenberghe et al., 2010) and our study therefore does not allow for a direct within-subjects comparison between FDG-PET and ^{18}F-flutemetamol. Other studies in AD, MCI, FTLD and controls have evaluated the degree of concordance between ^{11}C-PIB and FDG-PET at the group level (Furst et al., 2012) and also with regards to individual case classification, either on the basis of visual reads (Rabinovici et al., 2007) or semiquantitative retention levels (Devanand et al., 2010). In a substantial minority of patients with clinical FTLD, ^{11}C-PIB and FDG-PET visual reads were discordant (Rabinovici et al., 2007). Diagnostic performance in AD, MCI and controls appeared to be slightly better for ^{11}C-PIB than for FDG-PET, in particular in the MCI group (Devanand et al., 2010).

The prevalence of positive ^{18}F-flutemetamol scans in the elderly controls was similar to that observed in some of the other cohort studies (Devanand et al., 2010) and relatively low compared to most ^{11}C-PIB scanned cohorts (Aizenstein et al., 2008; Pike et al., 2011). This difference between cohorts will require further evaluation of ^{18}F-flutemetamol in larger samples distributed over a wider age range. Relatively small differences in age range could possibly have a large effect on the prevalence rates. Alternatively, difference in prevalence between cohorts may reflect differences in sensitivity between ligands or other methodological aspects. Thirdly, community-recruited cohorts are prone to selection bias which may lead to differences in prevalence of positive amyloid scans between cohorts.

The distance from the hyperplane as such cannot substitute for the retention values per se as a biomarker. It is a measure of the degree of confidence with which the classifier assigns the case to a given class and an expression of how robust this classification would remain e.g. with minor variations in study sample or hyperplane orientation. This 'confidence' may be related to (Ecker et al., 2010) but is theoretically and fundamentally distinct from the degree of severity of the underlying pathophysiological process. As a consequence, in our opinion, the distance to the hyperplane will not be useful directly as a biomarker, in contrast with the retention values per se.

A recent model of the progressive disease course of AD contends that an increase in β amyloid load precedes regional volume loss, prior to any clinical manifestations (Jack et al., 2010). Our cross-sectional study was not designed to test this model specifically. A longitudinal study design would be required and, moreover, our sample of healthy elderly controls was undersized to support the drawing of firm conclusions with respect to this model. In our elderly controls, more cases were classified as AD-like by the gray matter volume-based classifier than by the ^{18}F-flutemetamol-based classifier, differently

from what the above model would predict.

To conclude, the strength of this study is the application of two imaging modalities within the same subjects in a relatively large sample and the application of a mathematical procedure for diagnostic classification in the same manner for both modalities. A supervised machine learning technique can classify ^{18}F-flutemetamol scans in a binary way that is highly concordant with visual reads, corroborating the reliability of binary visual reads of ^{18}F-flutemetamol images (Vandenberghe et al., 2010). The gray matter volume-based classifier assigned more scans from healthy controls and more scans from MCI cases to the AD-like category than the ^{18}F-flutemetamol-based classifier, suggesting that overlap between healthy aging and AD may be more pronounced for MRI than for ^{18}F-flutemetamol.

Acknowledgements

The study was designed and sponsored by GE Healthcare. The PI RV and his coworkers had full access to the data, analysed the data and wrote the paper. RV is a senior clinical investigator of the Research Foundation Flanders (FWO), NN a postdoctoral fellow of the FWO. Funding support: Stichting Alzheimer Onderzoek grant 09013, 11020.

References

Abdulkadir, A., Mortamet, B., Vemuri, P., Jack, Jr, C. R., Krueger, G., Klöppel, S., 2011. Alzheimer's Disease Neuroimaging Initiative. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. Neuroimage 58, 785–92.

Aizenstein, H. J., Nebes, R. D., Saxton, J. A., Price, J. C., Mathis, C. A., Tsopelas, N. D., Ziolko, S. K., James, J. A., Snitz, B. E., Houck, P. R., Bi, W., Cohen, A. D., Lopresti, B. J., DeKosky, S. T., Halligan, E. M., Klunk, W. E., 2008. Frequent amyloid deposition without significant cognitive impairment among the elderly. Arch Neurol 65, 1509–1517.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Becker, J. A., Hedden, T., Carmasin, J., Maye, J., Rentz, D. M., Putcha, D., Fischl, B., Greve, D. N., Marshall, G. A., Salloway, S., Marks, D., Buckner, R. L., Sperling, R. A., Johnson, K. A., 2011. Amyloid-β associated cortical thinning in clinically normal elderly. Ann Neurol 69, 1032–1042.

Brilliant, M. J., Elble, R. J., Ghobrial, M., Struble, R. G., 1997. The distribution of amyloid beta protein deposition in the corpus striatum of patients with alzheimer's disease. Neuropathol Appl Neurobiol 23, 322–325.

Chételat, G., Villemagne, V. L., Bourgeat, P., Pike, K. E., Jones, G., Ames, D., Ellis, K. A., Szoeke, C., Martins, R. N., O'Keefe, G. J., Salvado, O., Masters, C. L., Rowe, C. C., , A. I. B., Group, L. R., 2010. Relationship between atrophy and beta-amyloid deposition in alzheimer disease. Ann Neurol 67, 317–324.

Chételat, G., Villemagne, V. L., Pike, K. E., Ellis, K. A., Bourgeat, P., Jones, G., O'Keefe, G. J., Salvado, O., Szoeke, C., Martins, R. N., Ames, D., Masters, C. L., Rowe, C. C.,

, A. I. B., of ageing (AIBL) Research Group, L. S., 2011. Independent contribution of temporal beta-amyloid deposition to memory decline in the pre-dementia phase of alzheimer's disease. Brain 134, 798–807.

Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti, R., Cerello, P., De Mitri, I., Retico, A., Nobili, F., Alzheimer's Disease Neuroimaging Initiative. 2011. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage 58, 469–80.

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

Devanand, D. P., Mikhno, A., Pelton, G. H., Cuasay, K., Pradhaban, G., Dileep Kumar, J. S., Upton, N., Lai, R., Gunn, R. N., Libri, V., Liu, X., van Heertum, R., Mann, J. J., Parsey, R. V., 2010. Pittsburgh compound b (11C-PIB) and fluorodeoxyglucose (18F-FDG) pet in patients with alzheimer disease, mild cognitive impairment, and healthy controls. J Geriatr Psychiatry Neurol 23, 185–198.

Duara, R., Loewenstein, D., Shen, Q., Barker, W., Potter, E., Varon, D., Heurlin, K., Vandenberghe, R., Buckley, C., 2012. ^{18}F-flutemetamol amyloid PET and structral MRI in the diagnosis of MCI and AD. Alzheimer's & Dementia in press.

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Brammer, M.J., Murphy, C., Murphy, D., the, M.R.C., Consortium, A.I.M.S., 2010. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. Neuroimage 49, 44–56.

Furst, A. J., Rabinovici, G. D., Rostomian, A. H., Steed, T., Alkalay, A., Racine, C., Miller, B. L., Jagust, W. J., 2012. Cognition, glucose metabolism and amyloid burden in alzheimer's disease. Neurobiol Aging 33, 215–225.

Gearing, M., Wilson, R. W., Unger, E. R., Shelton, E. R., Chan, H. W., Masters, C. L., Beyreuther, K., Mirra, S. S., 1993. Amyloid precursor protein (app) in the striatum in alzheimer's disease: an immunohistochemical study. J Neuropathol Exp Neurol 52 (1), 22–30.

Hodges, J. R., Davies, R. R., Xuereb, J. H., Casey, B., Broe, M., Bak, T. H., Kril, J. J., Halliday, G. M., 2004. Clinicopathological correlates in frontotemporal dementia. Ann Neurol 56, 399–406.

Hort, J., Bartos, A., Pirttil, T., Scheltens, P., 2010. Use of cerebrospinal fluid biomarkers in diagnosis of dementia across europe. Eur J Neurol 17, 90–96.

Ikonomovic, M., Klunk, W., Abrahamson, E., Mathis, C., Price, J., Tsopelas, N., Lopresti, B., Ziolko, S., Bi, W., Paljug, W., Debnath, M., Hope, C., Isanski, B., Hamilton, R., DeKosky, S., 2008. Post-mortem correlates of *in vivo* PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain 131, 1630–1645.

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

Klöppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack, Jr, C.R., Ashburne, J., Frackowiak, R.S. 2008a Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131, 2969–74.

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

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

Klunk, W., Wang, Y., Huang, G., Debnath, M., Holt, D., Shao, L., Hamilton, R., Ikonomovic, M., DeKosky, S., Mathis, C., 2003. The binding of 2-(4'-methylaminophenyl)benzothiazole to postmortem brain homogenates is dominated by the amyloid component. J Neurosci 23, 2086–2092.

Knopman, D. S., Boeve, B. F., Parisi, J. E., Dickson, D. W., Smith, G. E., Ivnik, R. J., Josephs, K. A., Petersen, R. C., 2005. Antemortem diagnosis of frontotemporal lobar degeneration. Ann Neurol 57, 480–488.

Koole, M., Lewis, D., Buckley, C., Nelissen, N., Vandenbulcke, M., Brooks, D., Vandenberghe, R., Van Laere, K., 2009. Whole-body distribution and radiation dosimetry of ^{18}F-GEO67: a radioligand for in vivo brain amyloid imaging. J Nucl Med 50, 818–822.

Leyton, C. E., Villemagne, V. L., Savage, S., Pike, K. E., Ballard, K. J., Piguet, O., Burrell, J. R., Rowe, C. C., Hodges, J. R., 2011. Subtypes of progressive aphasia: application of the international consensus criteria and validation using β-amyloid imaging. Brain 134, 3030–3043.

Liu, M., Zhand, D., Shen, D., 2012. Ensemble sparse classification of alzheimer's disease. Neuroimage 60, 1106–1116.

Lockhart, A., Lamb, J., Osredkar, T., Sue, L., Joyce, J., Ye, L., Libri, V., Leppert, D., Beach, T., 2007. PIB is a non-specific imaging marker of Aβ peptide-related cerebral amyloidosis. Brain 130, 2607–2615.

Lötjönen, J., Wolz, R., Koikkalainen, J., Julkunen, V., Thurfjell, L., Lundqvist, R., Waldemar, G., Soininen, H., Rueckert, D., , A. D. N. I., 2011. Fast and robust extraction of hippocampus from mr images for diagnostics of Alzheimer's disease. Neuroimage 56, 185–196.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 16, 187–198.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., Benali, H., 2009. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73–83.

Nelissen, N., Laere, K. V., Thurfjell, L., Owenius, R., Vandenbulcke, M., Koole, M., Bormans, G., Brooks, D., Vandenberghe, R., 2009. Phase I study of the PIB derivative ^{18}F-flutemetamol in healthy volunteers and patients with probable Alzheimer's disease. J Nucl Med 50, 1251–1259.

Nordberg, A., Carter, S., Rinne, J., Drzezga, A., Brooks, D., Vandenberghe, R., Perani, D., Forsberg, A., Langstrom, B., Scheinin, N., Karrasch, M., Nagren, K., Grimmer, T., Miederer, I., Edison, P., Okello, A., Van Laere, K., Nelissen, N., Vandenbulcke, M., Garibotto, V., Almkvist, O., Kalbe, E., Hinz, R., Herholz, K., 2012. A european multi-centre pet study of fibrillar amyloid in alzheimer's disease. European Journal of Nuclear Medicine and Molecular Imaging in press, in press.

Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A., 2012. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev. 36, 1140–52.

Padilla, P., López, M., Górriz, J.M., Ramírez, J., Salas-González, D., Álvarez, I.; Alzheimers Disease Neuroimaging Initiative, 2012. NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease. IEEE Trans Med Imaging 31, 207–16.

Pao, W. C., Dickson, D. W., Crook, J. E., Finch, N. A., Rademakers, R., Graff-Radford, N. R., 2011. Hippocampal sclerosis in the elderly: genetic and pathologic findings, some mimicking alzheimer disease clinically. Alzheimer Dis Assoc Disord 25, 364–368.

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

Pike, K. E., Ellis, K. A., Villemagne, V. L., Good, N., Chételat, G., Ames, D., Szoeke, C., Laws, S. M., Verdile, G., Martins, R. N., Masters, C. L., Rowe, C. C., 2011. Cognition and beta-amyloid in preclinical alzheimer's disease: data from the AIBL study. Neuropsychologia 49, 2384–2390.

Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50, 162–74.

Rabinovici, G. D., Furst, A. J., O'Neil, J. P., Racine, C. A., Mormino, E. C., Baker, S. L., Chetty, S., Patel, P., Pagliaro, T. A., Klunk, W. E., Mathis, C. A., Rosen, H. J., Miller, B. L., Jagust, W. J., 2007. 11C-PIB pet imaging in alzheimer disease and frontotemporal lobar degeneration. Neurology 68, 1205–1212.

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini, M.-L., Rosen, H., Prioleau-Latham, C. E., Lee, A., Kipps, C. M., Lillo, P., Piguet, O., Rohrer, J. D., Rossor, M. N., Warren, J. D., Fox, N. C., Galasko, D., Salmon, D. P., Black, S. E., Mesulam, M., Weintraub, S., Dickerson, B. C., Diehl-Schmid, J., Pasquier, F., Deramecourt, V., Lebert, F., Pijnenburg, Y., Chow, T. W., Manes, F., Grafman, J., Cappa, S. F., Freedman, M., Grossman, M., Miller, B. L., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134, 2456–2477.

Rudelli, R. D., Ambler, M. W., Wisniewski, H. M., 1984. Morphology and distribution of alzheimer neuritic (senile) and amyloid plaques in striatum and diencephalon. Acta Neuropathol 64, 273–281.

Suenaga, T., Hirano, A., Llena, J. F., Yen, S. H., Dickson, D. W., 1990. Modified bielschowsky stain and immunohistochemical studies on striatal plaques in alzheimer's disease. Acta Neuropathol 80, 280–286.

Thurfjell, L., Lundqvist, J. L. R., Koikkalainen, J., Soininen, H., Waldemar, G., Brooks, D., Vandenberghe, R., 2012. Combination of biomarkers: PET ^{18}F-flutemetamol imaging and structural MRI in dementia and mild cognitive impairment. Neurodegenerative Dis DOI:10;1159/000335381.

Vandenberghe, R., Van Laere, K., Ivanoiu, A., Salmon, E., Bastin, C., Triau, E., Hasselbalch, S., Law, I., Andersen, A., Korner, A., Minthon, L., Garraux, G., Nelissen, N., Bormans, G., Buckley, C., Owenius, R., Thurfjell, L., Farrar, G., Brooks, D., 2010. 18f-flutemetamol amyloid imaging in alzheimer disease and mild cognitive impairment: a phase 2 trial. Ann Neurol 68, 319–329.

Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., Boeve, B. F., Petersen, R. C., Jack, Jr, C. R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39, 1186–1197.