Psychiatry Research: Neuroimaging 212 (2013) 230–236

Contents lists available at SciVerse ScienceDirect

Psychiatry Research: Neuroimaging

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

Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI ☆

Juergen Dukart [a,b,e,*], Karsten Mueller [a], Henryk Barthel [b,c], Arno Villringer [a,b,d], Osama Sabri [b,c], Matthias Leopold Schroeter [a,b,d], for the Alzheimer's Disease Neuroimaging Initiative

[a] Max-Planck-Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany
[b] LIFE—Leipzig Research Center for Civilization Diseases, University of Leipzig, Germany
[c] Department of Nuclear Medicine, University of Leipzig, 04103 Leipzig, Germany
[d] Day Clinic of Cognitive Neurology, University of Leipzig, 04103 Leipzig, Germany
[e] LREN, Département des Neurosciences Cliniques, CHUV, Université de Lausanne, Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 28 July 2011
Received in revised form
4 April 2012
Accepted 20 April 2012

Keywords:
Multimodal imaging
Support vector machine classification
Multicenter validation
ADNI

ABSTRACT

The application of support vector machine classification (SVM) to combined information from magnetic resonance imaging (MRI) and [F18]fluorodeoxyglucose positron emission tomography (FDG-PET) has been shown to improve detection and differentiation of Alzheimer's disease dementia (AD) and frontotemporal lobar degeneration. To validate this approach for the most frequent dementia syndrome AD, and to test its applicability to multicenter data, we randomly extracted FDG-PET and MRI data of 28 AD patients and 28 healthy control subjects from the database provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and compared them to data of 21 patients with AD and 13 control subjects from our own Leipzig cohort. SVM classification using combined volume-of-interest information from FDG-PET and MRI based on comprehensive quantitative meta-analyses investigating dementia syndromes revealed a higher discrimination accuracy in comparison to single modality classification. For the ADNI dataset accuracy rates of up to 88% and for the Leipzig cohort of up to 100% were obtained. Classifiers trained on the ADNI data discriminated the Leipzig cohorts with an accuracy of 91%. In conclusion, our results suggest SVM classification based on quantitative meta-analyses of multicenter data as a valid method for individual AD diagnosis. Furthermore, combining imaging information from MRI and FDG-PET might substantially improve the accuracy of AD diagnosis.

1. Introduction

Imaging biomarkers such as regional atrophy as measured by structural magnetic resonance imaging (MRI), glucose hypometabolism as measured by [F18]fluorodeoxyglucose positron emission tomography (FDG-PET) or brain β-amyloid load as measured by PIB- (Pittsburgh compound B) and florbetaben-PET have been reported to be useful in diagnosis and/or differential diagnosis of dementia (Sabri et al., 1999, 2008; Hoffman et al., 2000; Rosen et al., 2002; Buckner et al., 2005; Diehl et al., 2004; Jeong et al., 2005; Diehl-Schmid et al., 2007; Edison et al., 2007; Fung and Stoeckel, 2007; Jack et al., 2008; Schroeter et al., 2007, 2008, 2009; Schroeter and Neumann, 2011; Davatzikos et al., 2008; Klöppel et al., 2007; Barthel et al., 2011). Therefore, it is now being suggested to incorporate such imaging markers into criteria for in vivo diagnosis of dementia (Dubois et al.,. 2007; Kipps et al., 2009).

Most previous biomarker studies focused on one specific biomarker or compared sensitivity and specificity of different single biomarkers (Fung and Stoeckel, 2007; Davatzikos et al., 2008; Habeck et al., 2008; Klöppel et al., 2007; Chaves et al., 2009; Habert et al., 2009; Horn et al., 2009; Ramirez et al., 2009a). Obviously, the combination of biomarkers potentially offers further improvements, and statistical methods such as multivariate pattern analyses using support vector machine (SVM) classifications not only enable automatic classification using one specific biomarker, but also provide a tool to combine two or more different biomarkers within the same classification model. For example, it has been shown that combining information from FDG-PET and MRI substantially improves detection (Hinrichs et al., 2009; Zhang et al., 2011) and differentiation of Alzheimer's disease dementia (AD)

☆ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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. ADNI investigators include (complete listing available at 〈http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf〉).

* Corresponding author at: Max-Planck-Institute for Human Cognitive and Brain Sciences, Stephanstr 1A 04103 Leipzig, Germany. Tel.: +49 341 99402413; fax: +49 341 99402221.
E-mail address: dukart@cbs.mpg.de (J. Dukart).

0925-4927/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pscychresns.2012.04.007

and frontotemporal lobar degeneration (Dukart et al., 2011). The volume-of-interest (VOI) approach used in Dukart et al. (2011), although less sensitive compared to whole-brain classification when using a single modality, was far superior to whole-brain classification when combined information from FDG-PET and MRI was used. VOIs used in this study were extracted from two comprehensive systematic and quantitative meta-analyses investigating both dementia syndromes in very large cohorts with anatomical likelihood estimates (ALE) (Schroeter et al., 2007, 2009). Accordingly, these VOIs represent the prototypical networks affected by these diseases and are not biased to a specific dataset. Furthermore, the preprocessing algorithm (Dukart et al., 2011) was designed to overcome difficulties which occur due to the use of different scanner types and different scanning sequences with different scaling and resolution. However, the generalizability of the results of that study (Dukart et al., 2011) was highly limited due to the low number of subjects and because data from only one center were used for SVM classification.

The goal of the present study was to further validate the approach proposed by Dukart et al. (2011) and to assess its generalizability to data from multicenter studies. To achieve this, we applied the identical preprocessing and classification algorithm to two different datasets. Classification accuracy results using FDG-PET and MRI data from the Clinic of Cognitive Neurology at the University of Leipzig were compared to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni-info.org). To avoid a classification bias towards the ADNI data (because substantially more control subjects and patients were available in this database than in the Leipzig cohort), we restricted the number of subjects and patients included from this database to make the numbers comparable with the Leipzig cohort. This is important because otherwise a combined classifier from both datasets would have mainly learned the distribution of subject and patient data in the ADNI cohort due to the substantially higher amount of subjects while practically ignoring the single center data. The ADNI database is a free access database containing, besides comprehensive neuropsychological and clinical evaluation, FDG-PET and MRI data of AD patients and healthy control subjects. We hypothesized that multicenter MRI and FDG-PET data obtained using different scanner types and sequences might be used to improve the accuracy of AD diagnosis in single clinical centers. Furthermore, we hypothesized that it would be possible to confirm our recent findings – namely improvement of individual dementia diagnosis with SVM classification of combined MRI and FDG-PET data and the aforementioned VOI approach based on meta-analyses.

2. Methods

2.1. Leipzig cohort

2.1.1. Subjects

We analyzed FDG-PET and T1-weighted MRI data of 21 patients (Table 1) with an early stage of probable AD, 14 patients with an early stage of FTLD and 13 control subjects. Patients were recruited from the Clinic of Cognitive Neurology at the University Hospital Leipzig. Probable AD was diagnosed according to the clinical NINCDS-ADRDA criteria (McKhann et al., 1984). Diagnosis of FTLD was based on clinical criteria suggested by Neary et al. (1998). The control group included subjects who visited the Clinic with subjective cognitive complaints, which were not objectively confirmed by a comprehensive neuropsychological and clinical evaluation. This control group was chosen because, in clinical practice, it is crucial to discriminate between these subjects and patients with an early stage of neurodegenerative disease. Informed consent was obtained from all subjects. The research protocol was approved by the Ethics Committee of the University of Leipzig, and was in accordance with the latest version of the Declaration of Helsinki.

2.1.2. MRI data

For each subject, a high-resolution T1-weighted MR scan was obtained consisting of 128 sagittal slices adjusted to the anterior commissure-posterior commissure (ACPC) line and with a slice thickness of 1.5 mm and pixel size of 1×1 mm^2. MRI was performed on two different 3T scanners (MedSpec 30/100, Bruker Biospin, Ettlingen Germany and Magnetom Trio, Siemens, Erlangen, Germany) using two different T1-weighted sequences (MDEFT or MP-RAGE with TR=1300 ms, TI=650 ms, TE=3.93 ms or TE=10 ms; FOV 25×25 cm^2; matrix=256×256 voxels). On the MedSpec scanner, only the MDEFT sequence and on the Magnetom Trio scanner, either MDEFT or MP-RAGE sequences were used.

2.1.3. FDG-PET data

Each subject also underwent FDG-PET imaging (~370 MBq) within a few weeks before or after the MR scan. All PET data were acquired on a Siemens ECAT EXACT HR+ scanner (CTI/Siemens, Knoxville, TN, USA) under a standard resting condition in 2-dimensional (2D) mode. Sixty-three slices were simultaneously collected with an axial resolution of 5 mm full width at half maximum (FWHM) and an in-plane resolution of 4.6 mm FWHM. After correction for attenuation, scatter, decay and scanner-specific dead-time, images were reconstructed by filtered back-projection using a Hann filter of 4.9 mm FWHM. The 63 transaxial slices obtained had a matrix of 128×128 voxels with an edge length of 2.45 mm. The resulting dynamic 0–60 min p.i. ECAT volume files were separated into single frames (ANALYZE format) using the import tool from the program MRIcro (http://www.sph.sc.edu/comd/rorden/mricro.html) and the last three frames of 10 min each, starting from 30 to 60 min post-injection, were included. Each subject's frames were spatially realigned to minimize inter-frame motion artefacts and a mean image of these three frames was calculated for each subject. These mean images were chosen for further analysis.

2.2. ADNI database

2.2.1. ADNI subjects

To validate the multimodal classification approach (Dukart et al., 2011) and to make it more reliable and generalizable to multicenter data, we extracted MR and FDG-PET images of 28 AD patients and 28 healthy control subjects (Table 1) from the ADNI database. The first available 28 AD patients were included into the study. Control subjects were selected to match AD patients for age and gender. This substantially smaller group size than currently available in the ADNI database was used for classification to insure a comparability between both cohorts in terms of group sizes. The ADNI is a partnership of the National Institute of Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations. Diagnosis of AD was based on NINCDS/ARDRA criteria (McKhann et al., 1984). Exclusion criteria for the ADNI data were any significant neurological disease other than AD, history of head trauma followed by persistent neurological deficits or structural brain abnormalities, psychotic features, agitation or behavioral problems within the last 3 months or history of alcohol or substance abuse. For most subjects multiple follow-up FDG-PET and MR scans are available. To ensure that our approach is applicable for the early diagnosis of dementia for all subjects, data from the first visit FDG-PET and MR scan were used.

2.2.2. ADNI MRI data

The MRI dataset included standard T1-weighted images obtained with different scanner types using volumetric MP-RAGE sequence varying in TR and TE with an in-plane resolution of 1.25×1.25 mm and 1.2 mm sagittal slice thickness. Only images obtained using 1.5 T scanners were used in this study. All images were preprocessed as described on the ADNI website (http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml) including distortion correction and B1 non-uniformity correction.

2.2.3. ADNI FDG-PET data

All ADNI subjects also underwent FDG-PET scanning obtained with different scanner types and using one of three different protocols: (1) dynamic: a 30-min, six-frame acquisition (six 5-min frames), with scanning from 30 to 60 min post-FDG injection; (2) static: a single-frame 30-min acquisition with scanning 30–60 min post-injection; and (3) quantitative: a 60-min dynamic protocol consisting of 33 frames, with scanning beginning at injection and continuing for 60 min that can be used to determine absolute glucose metabolic rate via kinetic modeling using an internal input function obtained from the internal carotid arteries. The majority of the scans in the ADNI study were acquired with the first acquisition protocol. The PET images further differed in resolution, image dimensions and count statistics. As with the FDG-PET data of the Leipzig cohort, the frames from 30 to 60 min post-injection were spatially realigned to minimize inter-frame motion artifacts and a mean image of these frames was calculated for each subject. These mean images were used for further analysis.

2.3. Preprocessing of MRI and FDG-PET data

For the MRI and FDG-PET data the preprocessing procedure as recently described in detail in Dukart et al. (2011) was applied. In short, this procedure included interpolation of both FDG-PET and MR images to an isotropic resolution

Table 1
Subject group characteristics.

	Leipzig cohort		ADNI dataset		ANOVA (d.f.,*F*,*P*)
	Controls	AD	Controls	AD	
Number	13	21	28	28	–
Male/female	7/6	9/12	20/8	19/9	–
Age (years)	53.9 ± 6.0	61.1 ± 6.7*	75.4 ± 4.6**	75.8 ± 7.2**	3,58.3, < 0.001
CDR (score)	0.23 ± 0.26	0.71 ± 0.25*	0.02 ± 0.09**	0.80 ± 0.25*	3,77.5, < 0.001
MMSE (score)	n.a.	23.2 ± 3.9	28.9 ± 1.3	23.8 ± 2.2*	2,41.0, < 0.001
Education (years)	12.3 ± 3.1	10.7 ± 3.1	16.5 ± 3.2	14.8 ± 3.5	3,14.9, < 0.001

Mean ± standard deviation. In the Leipzig cohort MMSE was missing for 1 AD patient.
AD—Alzheimer's disease, ADNI—Alzheimer's Disease Neuroimaging Initiative, CDR—Clinical Dementia Rating Scale, MMSE—Mini Mental State Examination, n.a.—not available.

* Significant difference to the control group from the same dataset,
** Significant difference to the same diagnosis group from the Leipzig cohort dataset.

of $1 \times 1 \times 1$ mm^3, bias correction for inhomogeneity artifacts for MR data, partial volume effect correction and masking of non-gray matter voxels in FDG-PET data and spatial normalization to an average size template created from all subjects using the DARTEL approach (Ashburner, 2007). The same deformations calculated based on an MRI template were applied to MRI and to coregistered FDG-PET images. After smoothing of FDG-PET and MR images with a Gaussian kernel of 12 mm FWHM, intensity normalization of FDG-PET data to the cerebellum (Dukart et al., 2010) and subsequent masking, an accurate anatomical overlap of both modalities was obtained. The binary mask was obtained by excluding all voxels in the first and the last template created by the DARTEL approach with a probability of below 0.2 for belonging to gray matter and using only voxels for classification that exceed this threshold in both templates. This mask was applied twice: firstly prior to smoothing to avoid misclassification, and secondly, after the smoothing to avoid big edge effects (Dukart et al., 2011). Both imaging modalities then have the same spatial orientation, resolution and effective smoothness. Individual intensity normalization to cerebellum also accounts for initially different count statistics of FDG-PET images. However, both imaging modalities are still in the space of the average template created from all subjects. To extract regional values corresponding to VOIs reported in the activation likelihood estimation (ALE) meta-analysis investigating AD (Schroeter et al., 2009) all images were normalized to Montreal Neurological Institute (MNI) space by applying affine-only transformations obtained when normalizing the average template from all subjects to the MNI space gray matter template provided by Statistical Parametric Mapping (SPM).

2.3.1. VOI extraction

VOI coordinates were extracted using the MRIcron 3D fill tool (http://www.sph.sc.edu/comd/rorden/mricron) as described in Dukart et al. (2011) from a comprehensive, systematic and quantitative ALE meta-analysis investigating biomarkers of AD in MR and FDG-PET images (Fig. 1, Table 2). These VOIs represent the maxima of atrophy or reductions in glucose utilization in AD.

The meta-analysis included a total number of 1351 AD patients and 1097 healthy control subjects (Schroeter et al., 2009). This meta-analysis extracted the prototypical network of AD by applying what is currently the most sophisticated and best-validated of coordinate-based voxel-wise meta-analyses, ALE. Because the coordinates in the meta-analysis were reported in the Talairach space, they were transformed to MNI space according to a formula proposed by Matthew Brett (published on the Internet: http://www.mrccbu.cam.ac.uk/Imaging/Common/mnispace.shtml). Six VOIs were extracted from each of the modalities by drawing a sphere with a radius of 5 mm around the reported coordinate and restricting the VOI to non-zero intensities within the sphere. The mean value of each VOI was used for classification.

2.3.2. SVM

SVM classification was conducted with the freely available LIBSVM software (Chang and Lin, 2001) using the Matlab interface. Multivariate pattern classification was performed with a linear kernel by identifying a separating linear hyperplane that maximizes the distance between different clinical groups based on VOI information. Additionally, the cost parameter c was optimized for each dataset by maximizing the leave-one-out cross-validation accuracy within the training cohort using a grid search algorithm by exponentially increasing the c and then refining the search grid around the detected accuracy maximum. The classifier trained using the obtained c was then applied to predict the independent cohort. The cross-validation of the trained SVM was performed by using the leave-one-out method. The procedure iteratively leaves out the information of each subject and trains the model on the remaining subjects for subsequent class assignment of the person that was not included in the training procedure. This validation method enables the generalization of the trained SVM to data that have never been presented to the SVM algorithms previously. The reported accuracy is the percentage of subjects correctly assigned to the clinical diagnosis. To improve the validity of this approach, separate classifiers were trained based either on the dataset from the ADNI database, the dataset of AD patients and control subjects from the Leipzig cohort or on combined data from both samples. Leaving-one-out cross-validation was performed for all classifiers. Furthermore, classifiers trained only on the ADNI cohort were applied to the Leipzig cohort and vice versa. SVM classification was applied separately to VOIs extracted from FDG-PET and MR data and to combined information from both imaging modalities. Feature combination was achieved by concatenating FDG-PET and MR features in a single vector.

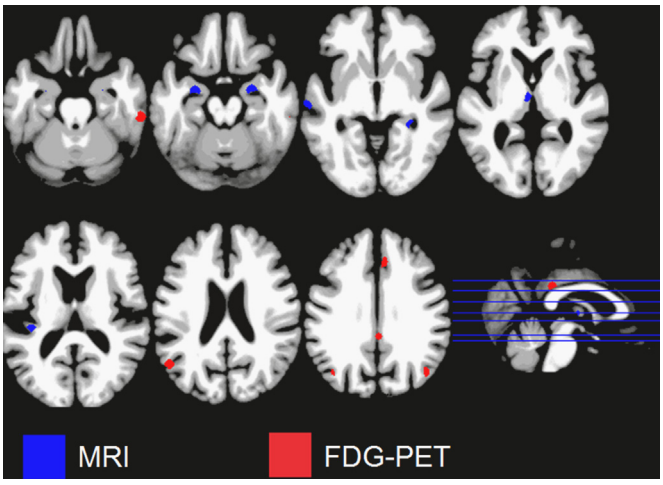

Fig. 1. VOIs used for differentiation between AD patients and control subjects plotted onto an average size brain. In blue MRI VOIs and in red FDG-PET VOIs are shown. Neurological convention (left is left), AD—Alzheimer's disease, FDG-PET—[F18]fluorodeoxyglucose positron emission tomography, MRI magnetic resonance imaging, VOI—volume-of-interest.

2.4. Statistical analysis

Group comparisons for age, education, the MMSE (Mini Mental State Examination; Folstein et al., 1975) and the CDR (Clinical Dementia Rating Scale; Morris, 1993) were performed by conducting one-way ANOVAs (analyses of variance). If an ANOVA revealed significant between-group differences, post-hoc Bonferroni t-tests were performed with a significance threshold of $p < 0.05$ corrected for multiple comparisons. Group differences regarding sex were evaluated using a chi-square test for independent samples. The statistical analysis was performed with the commercial software package SPSS 17.0 (http://www.spss.com/statistics/).

3. Results

3.1. Clinical characteristics

Clinical characteristics are illustrated in Table 1. The chi-square test for independent samples did not reveal any statistical differences in sex between the four groups [$\chi^2(3)=5.76$; $p=0.12$]. The ANOVAs revealed significant between-group differences in age,

Table 2
Coordinates of VOIs used for SVM classification.

AD vs. Controls						
FDG-PET	BA	Lat	*x*	*y*	*z*	SVM weight
Angular gyrus	39	L	−38	−68	37	2.4
Angular gyrus	39	R	43	−68	33	1.7
Posterior superior temporal sulcus	21/22					
Anterior medial frontal cortex	9/10	R	1	31	31	−2.9
Pregenual anterior cingulate gyrus	32					
Inferior precuneus	31	R	1	−36	27	6.2
Dorsal posterior cingulate cortex	23					
Posterior superior temporal sulcus	21/22	L	−51	−61	23	−0.7
Middle inferior temporal sulcus	20/21	R	59	−31	−23	2.2
MRI	BA	Lat	x	y	z	
Posterior insula	13	L	−38	−25	15	−0.2
Medial thalamus		L	−5	−13	3	−0.3
Hippocampal body/tail		R	31	−38	−6	4.1
Middle temporal gyrus/superior temporal sulcus	21/22	L	−63	−21	−5	1.5
Amygdala, anterior hippocampal formation, uncus, (trans-) entorhinal area	28/34	R	25	−8	−18	3.2
Amygdala, anterior hippocampal formation, uncus, (trans-) entorhinal area	28/34	L	−26	−8	−18	2.4

Coordinates are in MNI space (L left, R right). The absolute value of the weight (arbitrary units) indicates the importance of the corresponding region for separation between AD and control subjects relative to other VOIs.
AD—Alzheimer's disease, BA—Brodmann area, FDG-PET—[F18]fluorodeoxyglucose positron emission tomography, MRI—magnetic resonance imaging, VOI—volume-of-interest, SVM—support vector machine.

Table 3
Accuracy rates for VOI-based SVM classification for FDG-PET and MRI separately and for combined information.

	FDG-PET (%)	MRI (%)	FDG-PET & MRI (%)
ADNI dataset	87.5	80.4	85.7
Leipzig cohort	97.1	88.2	100.0
Combined dataset (ADNI and Leipzig cohort data)	91.1	80.0	90.0
ADNI (combined dataset)	87.5	78.6	83.9
Leipzig cohort (combined dataset)	97.1	82.4	100.0

Accuracy represents the percentage of subjects correctly assigned to the respective groups.
AD—Alzheimer's disease, ADNI—Alzheimer's Disease Neuroimaging Initiative, FDG-PET [F18]fluorodeoxyglucose positron emission tomography, MRI—magnetic resonance imaging, VOI—volume-of-interest, SVM—support vector machine.

CDR, MMSE and education. AD subjects in the ADNI dataset had, as expected, significantly lower MMSE scores compared with the ADNI control group [$t(54)=10.7$; $p<0.001$]. There was no significant difference in the MMSE between AD patients in the ADNI and the Leipzig cohort dataset [$t(46)=0.72$; $p=1.0$]. CDR scores also differed significantly between AD patients and control subjects in the ADNI [$t(54)=15.63$; $p<0.001$] and in the Leipzig cohort dataset [$t(32)=5.36$; $p<0.001$] but not between both groups of AD patients [$t(47)=1.23$; $p=0.93$]. The control group from the Leipzig cohort had significantly higher CDR scores in comparison to the control group from the ADNI database [$t(39)=3.87$; $p=0.025$]. Age was substantially different between the two datasets, with ADNI patients [$t(47)=7.27$; $p<0.001$] and ADNI control subjects [$t(39)=12.59$; $p<0.001$] being older than the corresponding group from the Leipzig cohort. Moreover, there was a significant difference between AD patients and control subjects in the Leipzig cohort set with regard to mean age [$t(32)=3.18$; $p=0.008$] but not in the ADNI dataset [$t(54)=0.22$; $p=1.0$].

3.2. SVM results

The differentiation accuracy for single modality classification using leaving-one-out cross-validation was highest with 97% using FDG-PET VOI information in the Leipzig cohort dataset (Table 3). Lowest accuracy with 80% was obtained using MRI VOI information from the ADNI database. Classification accuracy using FDG-PET information was far superior to MRI-based classification in both groups and similar to the accuracy of combined FDG-PET and MRI data. For this combined information, accuracy rates ranged between 86% and 100% with lowest accuracy using only the ADNI dataset and highest accuracy using the Leipzig cohort dataset. The overall accuracy for the combined dataset using both modalities was 90%, with sensitivity of 89% and specificity of 82% for subjects from the ADNI database and 100% sensitivity and specificity for data from the Leipzig cohort (Table 4).

Table 4
Differentiation rates for combined VOI information from FDG-PET and MRI.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	BER (%)
ADNI dataset	85.7	89.3	82.1	14.3
Leipzig cohort	100.0	100.0	100.0	0.00
Combined dataset	90.0	91.8	87.8	10.2

ADNI—Alzheimer's Disease Neuroimaging Initiative, BER—balanced error rate, FDG-PET—[F18]fluorodeoxyglucose positron emission tomography, MRI—magnetic resonance imaging, VOI—volume-of-interest.

One might assume that classifiers obtained from multicenter data as represented in the ADNI dataset enable a high discrimination accuracy for single center data. Indeed, the discrimination for Leipzig cohort patients using combined FDG-PET and MRI information from the ADNI dataset for training was very high with 97% whereas the opposite comparison predicting ADNI data using the Leipzig cohort dataset revealed lower, however still high, accuracy of 82% (Table 5). Accuracies observed using only FDG-PET or MRI were equal or lower compared to the use of combined information.

4. Discussion

In our study we investigated classification accuracies for detection of AD using FDG-PET, MRI or combined information from both imaging modalities in two independent cohorts. Thereby, we used VOIs centered to coordinates reported in a comprehensive meta-

Table 5
Accuracy rates for VOI-based SVM classification using independent cohorts.

	FDG-PET (%)	MRI (%)	FDG-PET & MRI (%)
ADNI→Leipzig cohort			
Accuracy	97.1	70.6	97.1
Sensitivity	100.0	81.0	100.0
Specificity	92.3	53.9	92.3
BER	3.8	32.6	3.8
Leipzig cohort→ADNI			
Accuracy	75.0	73.2	82.1
Sensitivity	82.1	85.7	78.6
Specificity	67.9	60.7	85.7
BER	25.0	26.8	17.9

ADNI—Alzheimer's Disease Neuroimaging Initiative, BER—balanced error rate, FDG-PET—[F18]fluorodeoxyglucose positron emission tomography, MRI—magnetic resonance imaging, SVM—support vector machine, VOI—volume-of-interest, → indicates the direction of prediction for each classification.

analysis investigating AD (Schroeter et al., 2009) as features for SVM classification. Very high accuracies for discrimination between AD patients and control subjects were obtained in both cohorts using only FDG-PET or combined information from both imaging modalities while MRI based accuracies were substantially lower for all comparisons. Combined information from FDG-PET and MRI were superior to both single modalities when using classifiers trained on one cohort for prediction of the second independent cohort. Our results are consistent with previous research showing that application of multivariate statistical methods on imaging markers might provide a substantial gain in accuracy for early detection and differentiation of dementia syndromes (Fung and Stoeckel, 2007; Davatzikos et al., 2008; Klöppel et al., 2007; Chaves et al., 2009; Hinrichs et al., 2009).

While all studies applying multivariate statistical methods have shown high detection and differentiation rates for different dementia syndromes, some of these increased accuracy rates have been obtained with a trade-off in generalizability of the proposed approaches to new datasets. This is the case because the number of features used to obtain high classification accuracies was determined by potentially biased feature selection methods. Thereby, a minimum number of regions are identified that provide optimum separation between different groups within the specific dataset. Subsequently, the same dataset is used to validate the classification approach based on selected features (Fung and Stoeckel, 2007; Davatzikos et al., 2008; Gerardin et al., 2009). This method, although providing optimum classification accuracy in a specific sample, does not necessarily enable high classification accuracy for data from other clinical centers as some of the features might be specific to the study cohort. Here we avoided this potential bias by using meta-analysis based VOIs and additionally validated our approach by the use of two independent cohorts. The obtained accuracies of up to 100% for the Leipzig cohort and of up to 87% for the ADNI dataset are very high. The results using only the ADNI cohort are comparable to previously reported accuracies using similar datasets for both FDG-PET and MRI (Hinrichs et al., 2009; McEvoy et al., 2009).

The results of our study are in line with previous findings suggesting that combining information from different imaging modalities which have been reported to be sensitive biomarkers for a specific neurodegenerative disorder can substantially improve detection and differentiation of dementia in comparison to single modality approaches (Fan et al., 2008; Dukart et al., 2011). However, this improvement combining multimodal information has so far only been observed if VOI information were used but not using whole-brain information from FDG-PET and MRI (Dukart et al., 2011). Our results indicate that this VOI approach is applicable to new data sets and does not depend on specific scanner types or sequences. The combination of VOI information from FDG-PET and MRI was superior or at least comparable to the best single-modality VOI-based classification for all comparisons. We suggest that regions extracted only from the AD ALE meta-analysis (Schroeter et al., 2009) might be sufficient to obtain a reasonable differentiation accuracy for AD patients and control subjects without any further feature selection procedures. Additionally, our results suggest that classifiers trained on data provided by ADNI, which is an open access and multicenter database, result in very good discrimination accuracy for new data from a single clinical center and so even increase the potential applicability of the proposed approach in the clinical environment.

The proposed approach can be easily implemented in a clinical environment using, for example, ADNI data to train a SVM classifier (e.g., with the svmtrain function implemented in LIBSVM, Chang and Lin, 2001) using the reported feature coordinates. It is important to note that the pre-processing algorithm should be kept constant between training and testing as the feature weights distribution might depend on some pre-processing steps. The obtained classifier can then be applied to estimate the likelihood of AD (or frontotemporal lobar degeneration as shown in our previous study, Dukart et al., 2011) in new subjects/patients based on their FDG-PET and MRI data. Thereby, the trained SVM model is applied on the mean values extracted from the individual subject's images at reported MNI coordinates using commonly available SVM implementations (e.g., the svmpredict function implemented in LIBSVM). The advantage of this approach is that additionally to the binary predicted label it also provides a probability value for the subject to belong to a specific category, thereby improving single subject diagnosis of AD as a supportive feature. Additionally, with a comparably low amount of features it enables a very high accuracy of 88–100% for discrimination of AD patients and control subjects which has now been validated using two independent cohorts. Additionally, as shown in a previous study, the proposed approach also provides a very good accuracy of up to 94% for differentiation of AD patients from patients with frontotemporal lobar degeneration (Dukart et al., 2011), which is another common type of neurodegenerative disease in aging. This additional sensitivity for both dementia syndromes is obtained by simple inclusion of features reported in meta-analyses discriminating patients with frontotemporal lobar degeneration from healthy control subjects (Schroeter et al., 2007, 2008). The accuracy of up to 92% obtained using the proposed algorithm for the three-group differentiation of AD patients, patients with frontotemporal lobar degeneration and healthy control subjects is the highest reported up to now, supporting the implementation of the proposed algorithm in clinical routine.

In our study we used only around 30% of available control subjects and AD patients from the ADNI database. Therewith, we wanted to avoid substantial differences in group sizes between the ADNI and Leipzig cohorts as it has been shown in recent research that increasing the group size of the training cohort also increases classification accuracies for discrimination of AD patients and control subjects (Abdulkadir et al., 2011). According to this study, we would expect even a further increase in classification accuracies using all data provided by the ADNI for training of the SVM classifier.

Some further limitations need to be considered regarding the interpretation of our results. Firstly, as the control group from the Leipzig cohort consists of subjects with subjective cognitive complaints, it is more likely to contain subjects with silent neurodegenerative disease compared to a population of non-complainers therewith potentially biasing the obtained accuracies. However, we would expect that a potential bias towards a silent neurodegenerative disease in the control group would rather lower the discrimination accuracy between control subjects and AD patients. Secondly, a further potential bias in the Leipzig cohort is the slightly

lower age of control subjects compared to AD patients, which might have contributed to the observed very high classification accuracy. Thirdly, our AD groups consisted of subjects who were at an already at a already manifest dementia stage of probable AD. For a higher clinical value of the proposed algorithm, it is therefore necessary to demonstrate that the proposed algorithm is also able to detect AD on even earlier stages such as mild cognitive impairment ahead of full manifestation of clinical symptoms. Lastly, a potential bias in differences in classification accuracies observed in both cohorts might be introduced by the different field strength of the acquired MRI data as all MRI scans used in the Leipzig cohort were acquired on a 3 T scanner while all images used from the ADNI cohort were acquired on a 1.5 T scanner. We think that for various reasons this limitation is unlikely to have introduced a substantial bias to our results as the obtained accuracies using only MRI information in both cohorts are comparable to each other and the observed differences are equally expressed in the FDG-PET modality.

4.1. Conclusion and perspectives

We investigated the applicability of the recently proposed approach for detection and differentiation of dementia syndromes using multimodal information from FDG-PET and MRI and meta-analysis based definition of VOIs to data from multiple clinical centers. For this purpose we used data from the ADNI database. The results are in line with previous findings that combining FDG-PET and MRI information improves detection of patients with early to moderate stages of AD. Furthermore, this approach allows using data from open access databases to improve clinical diagnosis of dementia in single clinical centers. Therefore, it has a high relevance for general clinical application. Additionally, our approach provides an easy method to integrate further biomarkers (such as from cerebrospinal fluid or other imaging modalities) to improve clinical diagnosis of various dementia syndromes. We conclude that SVM classification is a valid method for individual detection of frequent dementia syndromes using prototypical disease-related networks as extracted from meta-analyses and by combining multimodal imaging information.

Acknowledgment

Juergen Dukart, Henryk Barthel, Arno Villringer, Osama Sabri and Matthias L. Schroeter are supported by the Leipzig Research Center for Civilization Diseases (LIFE) at the University of Leipzig. LIFE is funded by the European Union, the European Regional Development Fund (ERFD) and the Free State of Saxony within the framework of the excellence initiative. Matthias L. Schroeter is also supported by the German consortium for frontotemporal lobar degeneration, funded by the German Federal Ministry of Education and Research.

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 (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

Abdulkadir, A., Mortamet, B., Vemuri, P., Jack, C.R., Krueger, G., Klöppel, S., Initiative, A. s.D.N., 2011. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. Neuroimage 58, 785–792.

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

Barthel, H., Gertz, H.J., Dresel, S., Peters, O., Bartenstein, P., Buerger, K., Hiemeyer, F., Wittemer-Rump, S.M., Seibyl, J., Reininger, C., Sabri, O., Florbetaben Study Group, 2011. Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. Lancet Neurology 10, 424–435.

Buckner, R.L., Snyder, A.Z., Shannon, B.J., LaRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C., Mintun, M.A., 2005. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. Journal of Neuroscience 25, 7709–7717.

Chang, C.C., Lin, C.J., 2001. LIBSVM: a Library for Support Vector Machines. Software available at: 〈http://www.csie.ntu.edu.tw/~cjlin/libsvm/〉.

Chaves, R., Ramírez, J., Górriz, J.M., López, M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., 2009. SVM-based computer-aided diagnosis of the Alzheimer's disease using *t*-test NMSE feature selection with feature correlation weighting. Neuroscience Letters 461, 293–297.

Davatzikos, C., Resnick, S.M., Wu, X., Parmpi, P., Clark, C.M., 2008. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41, 1220–1227.

Diehl, J., Grimmer, T., Drzezga, A., Riemenschneider, M., Förstl, H., Kurz, A., 2004. Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. Neurobiology of Aging 25, 1051–1056.

Diehl-Schmid, J., Grimmer, T., Drzezga, A., Bornschein, S., Riemenschneider, M., Förstl, H., Schwaiger, M., Kurz, A., 2007. Decline of cerebral glucose metabolism in frontotemporal dementia: a longitudinal 18F-FDG-PET-study. Neurobiology of Aging 28, 42–50.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurology 6, 734–746.

Dukart, J., Mueller, K., Horstmann, A., Vogt, B., Frisch, S., Barthel, H., Becker, G., Möller, H.E., Villringer, A., Sabri, O., Schroeter, M.L., 2010. Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. Neuroimage 49, 1490–1495.

Dukart, J., Mueller, K., Horstmann, A., Barthel, H., Möller, H.E., Villringer, A., Sabri, O., Schroeter, M.L., 2011. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. PLoSOne 6 (3), e18111. http://dx.doi.org/10.1371/journal.pone.0018111.

Edison, P., Archer, H.A., Hinz, R., Hammers, A., Pavese, N., Tai, Y.F., Hotton, G., Cutler, D., Fox, N., Kennedy, A., Rossor, M., Brooks, D.J., 2007. Amyloid, hypometabolism, and cognition in Alzheimer disease. An [11C]PIB and [18F]FDG PET study. Neurology 68, 501–508.

Fung, G., Stoeckel, J., 2007. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. Knowledge and Information Systems 11, 243–258.

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., Eustache, F., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, 2009. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47, 1476–1486.

Habeck, C., Foster, N.L., Perneczky, R., Kurz, A., Alexopoulos, P., Koeppe, R.A., Drzezga, A., Stern, Y., 2008. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. Neuroimage 40, 1503–1515.

Habert, M.O., Horn, J.F., Sarazin, M., Lotterie, J.A., Puel, M., Onen, F., Zanca, M., Portet, F., Touchon, J., Verny, M., Mahieux, F., Giron, A., Fertil, B., Dubois, B., 2009. Brain perfusion SPECT with an automated quantitative tool can identify prodromal Alzheimer's disease among patients with mild cognitive impairment. Neurobiology of Aging 32 (1), 15–23.

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

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. Journal of Nuclear Medicine 41, 1920–1928.

Horn, J.F., Habert, M.O., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., Giron, A., Fertil, B., 2009. Differential automatic diagnosis between Alzheimer's disease and

frontotemporal dementia based on perfusion SPECT images. Artificial Intelligence in Medicine 47, 147–158.
Jack Jr, C.R., Lowe, V.J., Senjem, M.L., Weigand, S.D., Kemp, B.J., Shiung, M.M., Knopman, D.S., Boeve, B.F., Klunk, W.E., Mathis, C.A., Petersen, R.C., 2008. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 131, 665–680.
Jeong, Y., Cho, S.S., Park, J.M., Kang, S.J., Lee, J.S., Kang, E., Na, D.L., Kim, S.E., 2005. 18F-FDG PET findings in frontotemporal dementia: an SPM analysis of 29 patients. Journal of Nuclear Medicine 46, 233–239.
Kipps, C.M., Hodges, J.R., Fryer, T.D., Nestor, P.J., 2009. Combined magnetic resonance imaging and positron emission tomography brain imaging in behavioural variant frontotemporal degeneration: refining the clinical phenotype. Brain 132, 2566–2578.
Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fos, N.C., Jack Jr, C.R., Ashburner, J., Frackowiak, R.S., 2007. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.
McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., 2009. Initiative, A. s. D. N., 2009. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 251, 195–205.
McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944.
Morris, J.C., 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43, 2412–2414.
Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., Boone, K., Miller, B.L., Cummings, J., Benson, D.F., 1998. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 51, 1546–1554.
Ramirez, J., Gorriz, J.M., Salas-Gonzales, D., Romero, A., Lopez, M., Alvarez, I., Gomez-Rio, M., 2009a. Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. Information Sciences http://dx.doi.org/10.1016/j.ins.2009.05.012.
Rosen, H.J., Gorno-Tempini, M.L., Goldman, W.P., Perry, R.J., Schuff, N., Weiner, M., Feiwell, R., Kramer, J.H., Miller, B.L., 2002. Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology 58, 198–208.
Sabri, O., Ringelstein, E.B., Hellwig, D., Schneider, R., Schreckenberger, M., Kaiser, H.J., Mull, M., Buell, U., 1999. Neuropsychological impairment correlates with hypoperfusion and hypometabolism but not with severity of white matter lesions on MRI in patients with cerebral microangiopathy. Stroke 30, 556–566.
Sabri, O., Kendziorra, K., Wolf, H., Gertz, H.J., Brust, P., 2008. Acetylcholine receptors in dementia and mild cognitive impairment. European Journal of Nuclear Medicine and Molecular Imaging 35 (Suppl. 1), 30–45.
Schroeter, M.L., Raczka, K., Neumann, J., Yves von Cramon, D., 2007. Towards a nosology for frontotemporal lobar degenerations-a meta-analysis involving 267 subjects. Neuroimage 36, 497–510.
Schroeter, M.L., Raczka, K., Neumann, J., von Cramon, D.Y., 2008. Neural networks in frontotemporal dementia—a meta-analysis. Neurobiology of Aging 29, 418–426.
Schroeter, M.L., Stein, T., Maslowski, N., Neumann, J., 2009. Neural correlates of Alzheimer's disease and mild cognitive impairment: a systematic and quantitative meta-analysis involving 1351 patients. Neuroimage 47, 1196–1206.
Schroeter, M.L., Neumann, J., 2011. Combined imaging markers dissociate Alzheimer's disease and frontotemporal lobar degeneration—an ALE meta-analysis. Frontiers in Aging Neuroscience 3, 10.
Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55, 856–867.