



Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease



I. Beheshti*, H. Demirel, for the Alzheimer's Disease Neuroimaging Initiative

Biomedical Image Processing Lab, Department of Electrical & Electronic Engineering, Eastern Mediterranean University, Gazimagusa, Mersin 10, Turkey

ARTICLE INFO

Article history:

Received 12 March 2015

Accepted 10 July 2015

Keywords:

Alzheimer's disease
Voxel-based morphometry
Probability distribution function
Structural MRI
Statistical feature extraction
Computer-aided diagnosis
Classification
Fisher criterion

ABSTRACT

High-dimensional classification methods have been a major target of machine learning for the automatic classification of patients who suffer from Alzheimer's disease (AD). One major issue of automatic classification is the feature-selection method from high-dimensional data. In this paper, a novel approach for statistical feature reduction and selection in high-dimensional magnetic resonance imaging (MRI) data based on the probability distribution function (PDF) is introduced. To develop an automatic computer-aided diagnosis (CAD) technique, this research explores the statistical patterns extracted from structural MRI (sMRI) data on four systematic levels. First, global and local differences of gray matter in patients with AD compared to healthy controls (HCs) using the voxel-based morphometric (VBM) technique with 3-Tesla 3D T1-weighted MRI are investigated. Second, feature extraction based on the voxel clusters detected by VBM on sMRI and voxel values as volume of interest (VOI) is used. Third, a novel statistical feature-selection process is employed, utilizing the PDF of the VOI to represent statistical patterns of the respective high-dimensional sMRI sample. Finally, the proposed feature-selection method for early detection of AD with support vector machine (SVM) classifiers compared to other standard feature selection methods, such as partial least squares (PLS) techniques, is assessed. The performance of the proposed technique is evaluated using 130 AD and 130 HC MRI data from the ADNI dataset with 10-fold cross validation¹. The results show that the PDF-based feature selection approach is a reliable technique that is highly competitive with respect to the state-of-the-art techniques in classifying AD from high-dimensional sMRI samples.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In older adults, Alzheimer's disease (AD) is a brain disorder that gradually impairs regions of the brain that are responsible for memory, learning, and higher executive functioning [1,2]. Currently, the detection of AD is based on clinical examinations and assessments of perception and behavior as indicators emerging in the later disease stages. Neuroimaging measures of structural changes and functional activities in the brain may be a good method for early detection of AD. Among the several neuroimaging techniques used in AD diagnosis, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission

computed tomography (SPECT), MRI is more widely used because of its excellent spatial resolution with good tissue contrast [3] without the need for radioactive pharmaceutical injection, as is required with PET and SPECT [4,5]. Regional and global (whole-brain) atrophy measurements are provided via MRI, and atrophy measured on structural MRI is a powerful biomarker of the stage and intensity of the neurodegenerative aspect of AD pathology [6]. Several studies have used structural MRI feature extraction for classification. Some of these studies are based on morphometric methods [7–9], region of interest (ROI)/volume of interest (VOI) [10–12], and gray matter voxels in the automatic segmentation of images [13].

The aim of this study was to introduce a novel statistical feature selection method based on the probability distribution function (PDF) of the VOI, which can be considered a lower-dimensional feature vector representing sMRI images. The PDF is assumed to be the statistical pattern of the VOI representing the entire sMRI.

The dimensionality of the PDF-based feature vector can be adjusted by changing the number of bins of the PDF. The proposed PDF-based method not only extracts the selected statistical features but also reduces the dimensionality of the input vectors to feature vectors. The PDF-based feature vector calculation process

* Corresponding author. Tel.: +90 392 630 1301x1093; fax: +90 392 630 1648.

E-mail addresses: iman.beheshti@cc.emu.edu.tr (I. Beheshti), hasan.demirel@emu.edu.tr (H. Demirel).

¹ Data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). ADNI investigators other than those listed above contributed to study design, implementation or data provision but did not participate in the analyses or writing of this report. The complete listing of ADNI investigators is available at http://www.loni.ucla.edu/ADNI/Data/ADNI_Authorship_List.pdf.

does not require matrix operations, making the feature extraction process computationally cheaper compared to alternative dimensionality reduction methods such as partial least squares (PLS). In this context, it is apparent that the computational cost of PDF calculation is negligibly low when compared to PLS. The proposed work was accomplished using four steps to develop an automatic computer-aided diagnosis (CAD) technique for AD diagnosis. First, a statistical method was used based on the VBM technique plus Diffeomorphic Anatomical Registration using the Exponentiated Lie algebra (DARTEL) approach to analyze group-wise comparisons between a cross-sectional structural MRI scans diseased group and normal controls [6,14,15]. Based on the VBM plus DARTEL approach, overall and regional structural gray matter alterations were investigated to define regions with a significant decline of gray matter in patients with AD compared to the healthy controls (HCs). Second, these specified areas (gray matter loss in AD patients) were employed as masks with the template and extracted voxel values from the VOI to form the raw feature vectors. These raw feature vectors went through further data reduction or selection processes before being used by the classifier. Third, a novel statistical feature vector generation using probability distribution functions (PDFs) extracted from the respective 3D mask regions of sMRI was used for classification. The PDF approach can help in two ways: (1) dimensionality reduction and (2) compressing the statistical information of the high-dimensional data into a lower-dimensional vector. PDF pattern recognition has been used successfully in a number of applications, including face recognition [16–18]. In addition, an automatic approach based on the Fisher criterion was used to determine the optimal number of bins of the histogram generating the PDF. This approach adaptively determines the number of PDF bins based on the training data in each fold instead of using a fixed one. Fourth, the performance of the proposed statistical feature-selection technique was evaluated using SVM classifiers.

The remainder of this paper is arranged as follows: Section 2 provides statistics of the data used in the work and Section 3 describes the methodology used to design an automatic CAD tool based on the PDF. The evaluation experiments and an analysis of the proposed method are described in Section 4, and the conclusion is drawn in Section 5.

2. Material

2.1. Image acquisition

MRI images and data used in this work were obtained from the 3 T MRI protocol of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). Briefly, the

protocol included T1-weighted MRI images based on a scanner by Siemens with acquisition plane=sagittal, acquisition type=3D, coil=PA, flip angle=9.0 degrees, matrix X/Y/Z=240.0/256/176 pixel, mfg model=Skyra, pixel spacing X/Y=1.0/1.0 mm, pulse sequence=GR/IR, slice thickness=1.2 mm, and TE/TI/TR=2.98/900/2300 ms.

2.2. Subjects

The group of patients with AD contained 130 people aged 57 to 91 years (mean 75.88 ± 7.54 years). The Mini Mental State Examination (MMSE) and Clinical Dementia Ratio (CDR) scores ranged from 10 to 28 (mean 22.33 ± 3.27) and 0.5 to 2 (mean 0.80 ± 0.37), respectively. The second group contained 130 HCs aged 56 to 88 years (mean 74.49 ± 6.13 years). The MMSE for this group ranged from 27 to 30 (mean 29.26 ± 0.80) and the CDR was zero. In a direct comparison between the HC and AD groups, there were no significant differences in age or the number of gender subjects.

3. Methodology of the CAD system

In this section, the methodology is presented based on the PDF approach to design an automatic CAD system for MRI classification. First, the VBM plus DARTEL approach process was used to perform pre-processing on 3D MRI data. Second, a feature-extraction method was employed based on VBM plus DARTEL analysis. Third, an adaptive PDF-based data-selection method was proposed, as a novel statistical data-selection mechanism representing the statistical pattern of VOI of high-dimensional sMRI data in a low-dimensional space. The dimension of the PDF-based vector depended directly on the number of bins used in the histogram of the VOI, which was then normalized into the PDF. The optimal number of bins was obtained by maximizing the Fisher criterion among the possible number of bins. Finally, to evaluate the proposed technique, classifiers such as SVM were used. Fig. 1 illustrates the framework of the proposed CAD system.

3.1. MRI data pre-processing

Data pre-processing was performed using Statistical Parameter Mapping (SPM) software version 8 (Wellcome Trust Centre for Neuroimaging, London, UK; available at: <http://www.fil.ion.ucl.ac.uk/spm>) and the voxel-based morphometry toolbox version 8 (<http://dbm.neuro.uni-jena.de/vbm>). VBM, introduced by Ashburner and Friston, is a method used to assess whole-brain structure with voxel-by-voxel comparisons, which has been developed to analyze tissue concentrations or volumes between subject groups to distinguish degenerative diseases with dementia [4,14]. Recently, VBM has

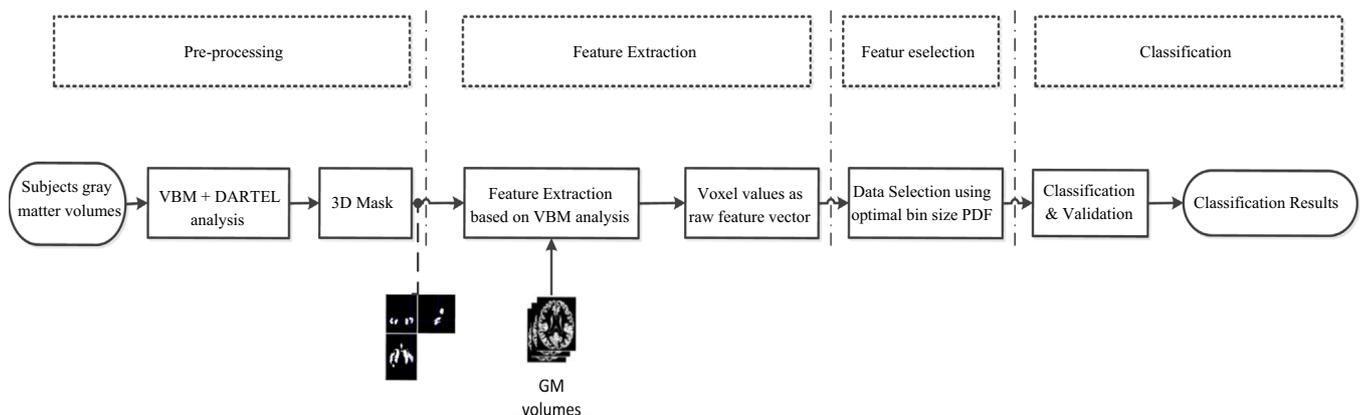


Fig. 1. The framework of proposed PDF-based CAD system classifying AD.

been applied to detect early atrophic changes in AD [3,19–21]. It can provide statistical results in comparisons of patients with AD to HCs [3,22]. To enhance inter-subject registration of the MRI images, DARTEL was applied [23,24], which has been found to optimize the sensitivity of such analyses by using the Levenberg–Marquardt strategy as compared to standard VBM [23,25]. Moreover, the DARTEL algorithm provides precise and accurate localization of structural damage on the MRI images [3,4]. In the VBM8 toolbox, registration to standard Montreal Neurological Institute (MNI) space is an important process, which contains linear affine transformation and nonlinear deformation by using high-dimensional DARTEL normalization. This process involves using the DARTEL template generated from 550 healthy control participants (defined by default settings of VBM8) [26]. The normalized segmented images were modulated by using a nonlinear deformation, which allows for comparing absolute amounts of tissue corrected for individual differences in brain size [26]. Finally, the segmented images were spatially smoothed with an 8 mm full-width-half-maximum (FWHM) Gaussian kernel. After spatial pre-processing, the smoothed, modulated, DARTEL warped and normalized gray matter datasets were used for statistical analysis. Regional gray matter volume changes were generated by voxel-based analysis over the whole brain. Fig. 2 illustrates the processing pipeline of the VBM analysis. To detect gray matter volume reductions in patients with AD, a two-sample t -test in SPM8 was used. Age was applied into the matrix design as a nuisance variable. To avoid possible edge effects between gray matter and white matter or cerebrospinal fluid (CSF), the absolute threshold masking was 0.1. Significance was set at a p -value of $\square 0.01$ with correction for family-wise error (FWE) and an extent threshold of 1400 adjacent voxels for two-sample comparisons. Between-group differences in demographics and clinical parameters among or between subgroups were executed by Statistical Package for Social Sciences software (SPSS version 16.0) by using an independent sample t -test, and $p \square 0.05$ was considered significant.

3.2. Feature extraction and data reduction and selection

A feature-extraction procedure based on VBM plus DARTEL analysis was applied to isolate the VOI. The regions of decreased gray matter volume obtained using VBM plus DARTEL analysis in patients who suffered from AD were segmented using a 3D mask. This mask was applied to the gray matter density volumes resulting from the VBM plus DARTEL analysis to extract voxel values as raw feature vectors. It is important to separate the data used for VBM 3D mask generation from the data used for classification. In other words, the data to model the 3D mask must explicitly come from the training set. In this context, we divided the dataset for VBM mask generation within each outer cross-validation fold separately. In other words, we randomly divided our subjects into 10 folds with the same number of AD and HC subjects in each fold. In each iteration, we used one fold for testing and 9 folds for training. Based on each training dataset, we performed VBM plus DARTEL analysis to reveal regions of decreased gray matter volume in patients as a 3D mask. In total, we defined 10 different masks with different lengths (e.g. from 59,395 to 69,170 voxels). The respective 3D masks were used in

the respective iteration to extract features from the training and testing datasets. The raw feature space in the VBM extracted feature set was very high in comparison to the number of samples. Because the sample feature vectors spanned a very small region in the feature vector space, data reduction was desired in post-processing. In this context, it is preferable to reduce the dimensionality of sMRI datasets. Therefore, the dimensionality of extracted raw feature vectors is reduced statistically by means of PLS and PDF.

3.2.1. Feature reduction based on PLS

PLS is a statistical algorithm for modeling the relationship between two datasets: $X \subset R^N$ and $Y \subset R^M$. Recently, the PLS data-reduction approach has been used successfully in a number of applications for machine-learning in AD [27–30]. After observing n data samples, PLS decomposes the $n \times N$ and the $n \times M$ matrices of zero mean variables X and Y , respectively, into the following form [27,31]:

$$X = TP^T + E$$

$$Y = UQ^T + F \quad (1)$$

where T and U are $n \times A$ matrices of the A extracted score vectors, P and Q are $N \times A$ and $M \times A$ matrices of loadings, and E and F are $n \times N$ and the $n \times M$ error matrices [27]. In this study, in each fold the PLS algorithm was applied to X (training dataset) and Y (training data label) in order to obtain score and loading matrices. In addition, a weight matrix was obtained from the training dataset to compute a score matrix for the testing dataset [27]. Next, score vectors obtained from the training and test datasets were used as feature vectors by SVM classifiers. Fig. 3 illustrates the pipeline of the PLS feature-reduction procedure.

3.2.2. Statistical feature selection based on PDF

The PDF of a raw feature vector extracted from VOI is a statistical description of the distribution of occurrence probabilities of voxel values that can be considered a feature vector representing a high-dimensional vector in a lower-dimensional space. In a mathematical sense, a PDF can be defined as a vector of probabilities representing the probability of the voxel values that fall into various disjointed intervals, known as bins. Given a raw vector extracted from VOI, the PDF, H , of the raw vector met the following conditions [16,18]:

$$H = [p_1, p_2, p_3, \dots, p_m], \quad p_i = \frac{\eta_i}{N}, \quad i = 1, 2, \dots, m \quad (2)$$

where η_i is the number of voxels falling into the i th bin, m is the number of bins, and N is the total number of voxels in the 3D mask. In the classification stage, the PDF, H , of raw vectors was used in the representation of the training and test data. The number of bins adjusts the dimensionality of a PDF vector. In this work, the number of bins was assumed to vary from 2 to 100.

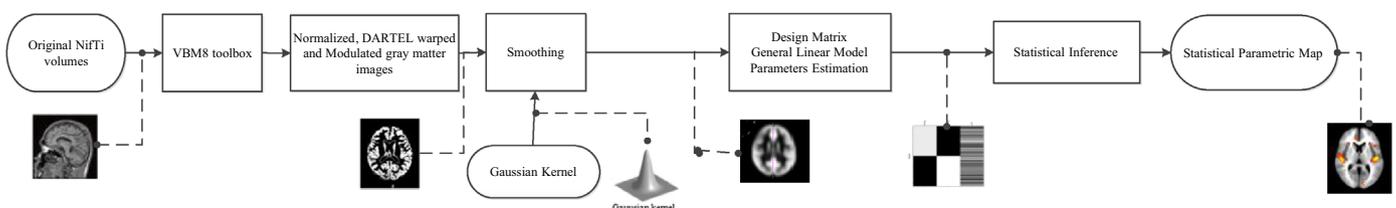


Fig. 2. The VBM plus DARTEL processing pipeline on sMRI data using SPM8.

3.2.3. Optimal number of bins based on Fisher criterion

To select the optimal number of bins, an automatic method was used, based on the Fisher criterion, $J(w)$, given in Eq. (3):

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (3)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix, respectively [32]. For the two classes, C_1 and C_2 , the between-class scatter and within-class scatter matrices are defined as:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (4)$$

$$S_W = \sum_{H_i \in C_1} (H_i - \mu_1)(H_i - \mu_1)^T + \sum_{H_i \in C_2} (H_i - \mu_2)(H_i - \mu_2)^T \quad (5)$$

where μ_1 is the mean of the PDF vectors in class 1 and μ_2 is the mean of the PDF vectors in class 2, and $w = S_W^{-1}(\mu_1 - \mu_2)$.

The main steps in the proposed algorithm are summarized in the pseudo code shown in algorithm 1. The number of bins (N_{bin}) of histogram H_i was iteratively incremented from 2 to 100, using a training set of each fold for calculating the respective Fisher criterion values. The optimal number of bins, N_{opt} , maximizing the Fisher criterion was selected to be used as the optimal dimension of the test and training data in each fold through the cross-validation process.

Algorithm 1. Optimal number of bins selection procedure.

```

1:  $V \leftarrow \text{component\_set}(\text{Data}_{\text{Train}}, \text{Label}_{\text{Train}})$ 
2: number of bin  $\leftarrow \emptyset, N_{bin} = 100$ 
3: for  $n=2$  to  $N_{bin}$  do
4:    $H_i \leftarrow \text{compute\_histogram}(X_i, n)$ 
5:    $(S_B, S_W) \leftarrow \text{compute\_scatter}(H_i, \text{Label}_{\text{Train}})$ 
6:    $\mu_1 \leftarrow \text{mean}(H_i \text{ class1})$ 
7:    $\mu_2 \leftarrow \text{mean}(H_i \text{ class2})$ 
8:    $w = S_W^{-1}(\mu_1 - \mu_2)$ 
9:    $\varphi(n) \leftarrow \frac{w^T S_B w}{w^T S_W w}$ 
10: end for
11:  $N_{opt} \leftarrow \arg \max_{n \in \{2, \dots, N_{bin}\}} \varphi(n)$ 

```

3.3. Classification implementation

To distinguish patients with AD from the HCs, the classification model in terms of the SVM algorithm was established [33].

3.3.1. The SVM

The SVM is a powerful classifier based on statistical learning principles. In several papers, SVM is used to correctly classify unseen patterns [9,34–36]. During SVM training, SVM maximizes the distance from patterns to the class-separating hyper-plane. Generally, the patterns are not linearly separable; therefore, non-linear kernel transformation is performed. There are various kernels that can be used during SVM training, including linear, quadratic, polynomial, and radial basis function (RBF) kernels. SVM was performed using the LIBSVM software package (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). In this work, SVM with linear and RBF kernels were used. The RBF model has two parameters that need to be selected, C (regularization) and γ (controls the kernel width), in which the performance of the classifier depends on these parameters. To evaluate the performance of the classifier, a procedure of two cross-validations (CVs) was combined with a grid search. This was done to elude unwarp bias in the estimation of accuracies produced by the CV procedure. This procedure

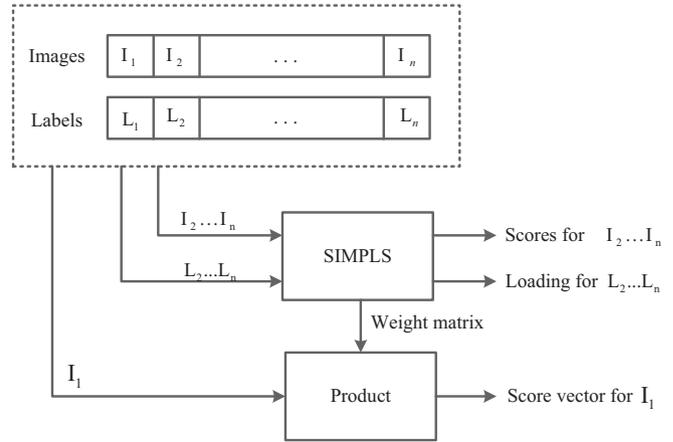


Fig. 3. Diagram of the PLS based feature extraction [27].

consisted of two nested loops. In the outer loop, the data were split into K_1 folds. At each step, one fold was used as a test and the remaining $K_1 - 1$ folds were used for training and validation. In the inner loop, the training data ($K_1 - 1$ folds) were further divided into K_2 folds. For each combination of C and γ , the classifier was trained using the training data and its performance was assessed using the fold left for validation by estimating the classification accuracy. One fold was left for validation and the remaining $K_2 - 1$ fold for training was combined with the grid search to determine the optimal parameters. In the grid search, the values of C and γ varied logarithmically from 2^{-5} to 2^{20} and from 2^{-15} to 2^{15} , respectively. The inner loop was repeated K_2 times and the accuracy of the classifier was obtained across the K_2 folds for every combination of C and γ . Optimal parameters were selected that produced maximum average accuracy across the K_2 folds. Then, the class label of the test data was predicted, which was left out in the outer loop using selected optimal parameters. The above procedure was repeated K_1 times by leaving a different fold as test data, which was used to compute the classification accuracy. For SVM with a linear kernel, only the C parameter was optimized. In this work, $K_1 = 10$ and $K_2 = 10$ were used. Fig. 4 illustrates the pipeline of the 10-fold cross-validation procedure.

3.3.2. Classification performance

The performance of a classifier is measured by using the accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC) based on 10-fold cross validation. These parameters are computed as follows [37]:

$$\text{ACC} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (6)$$

$$\text{SEN} = \text{TP} / (\text{TP} + \text{FN}) \quad (7)$$

$$\text{SPE} = \text{TN} / (\text{TN} + \text{FP}) \quad (8)$$

where TP, TN, FN, and FP are the number of true positives, true negatives, false negatives, and false positives, respectively. TP, TN, FN, and FP are determined as follows:

- TP: By counting the number of patients with AD correctly identified as AD.
- TN: By counting the number of HCs correctly identified as HCs.
- FN: By counting the number of patients with AD incorrectly identified as HCs.
- FP: By counting the number of HCs incorrectly identified as AD.

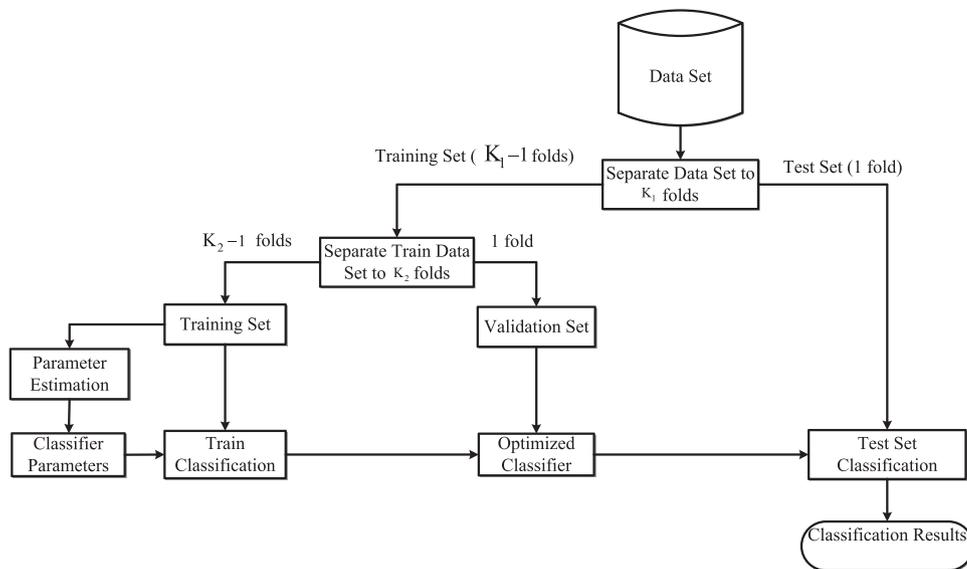


Fig. 4. Pipeline of the 10-fold cross validation procedure.

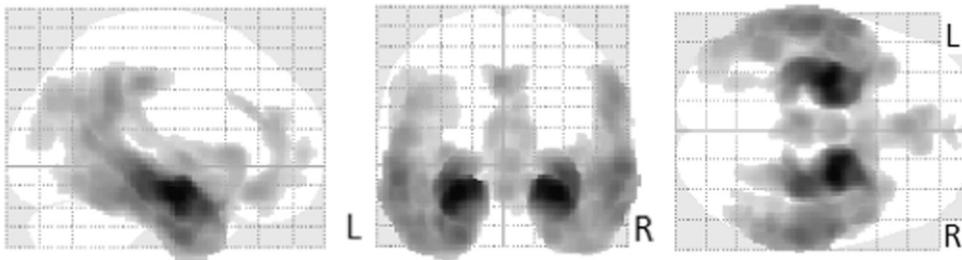


Fig. 5. Comparison of gray matter volume among 117 patients with AD and 117 HCs in fold 1 training by VBM using SPM8 plus DARTEL (FWE corrected at $p < 0.01$ and extend threshold $K = 1400$).

4. Experimental results and discussion

In this section, the experimental results of VBM plus DARTEL analysis on 3D MRI are reported to reveal the significance of the volumetric regions with atrophy in patients, contributing to VOI. The performance of the classification of AD using a 10-fold cross-validation is also presented for four cases: (1) performance of the raw features (VBM features) dataset, (2) performance of the PLS method, (3) performance of the proposed PDF technique, and (4) performance of the PDF technique using the optimal number of bins. Two types of SVM classifiers, namely SVM-linear and SVM-RBF, were used for AD classification. ACC (%), SEN (%), SPE (%), and AUC (%) performance metrics were used to assess the different scenarios.

4.1. Voxel-based morphometry on gray matter

VBM plus DARTEL revealed a significant decline of gray matter volume in the right hippocampus, left hippocampus, right inferior parietal lobe, and right anterior cingulate in patients with AD compared to the HCs. Fig. 5 shows the brain regions where there was significant atrophy in gray matter volume in AD patients compared to HCs in fold 1 training. The voxel locations of these significant regions were used as a 3D mask in each fold. This mask was applied to the gray matter density volume results from the segmentation step in the VBM plus DARTEL analysis to extract voxel values as raw feature vectors.

4.2. Performance of raw feature representation

The complete MRI dataset from the ADNI database consisted of 260 samples. Table 1 presents the ACC, SEN, SPE, and AUC obtained by 10-fold cross validation using SVM-linear and SVM-RBF classifiers for raw feature vectors obtained by masking after VBM plus DARTEL analysis.

4.3. Performance of PLS method

The feature reduction using PLS was accomplished by extracting raw feature data from VOI obtained from VBM analysis. The extracted raw feature vectors were reduced to lower-dimensional feature vectors of up to 100 components using PLS. Table 2(a) presents the ACC, SEN, SPE, and AUC obtained from 10-fold cross-validation for SVM classifiers for changing dimensionality. According to Table 2(a), it is clear that the maximum accuracy (90.76%) is yielded with SVM-RBF when the dimensionality is 80. The accuracy is 4.74% higher than the same classifier with all raw features used in Table 1. The rest of the results in Table 2(a) are also higher than the raw data for SEN, SPE, and AUC. The results reported in Tables 1 and 2(a) indicate that the PLS performance using SVM-linear and SVM-RBF classifiers is higher than with the raw data.

4.4. Performance of proposed PDF-based technique

The feature selection using PDF was accomplished by extracting raw feature data from VOI obtained using VBM analysis. The

extracted raw feature vectors were reduced to lower-dimensional feature vectors of up to 100 components by changing the number of bins of the PDF. Table 2(b) and Fig. 6 present the ACC, SEN, SPE, and AUC obtained by 10-fold cross-validation using SVM-linear and SVM-RBF classifiers. The results reported in Table 2(a) and (b) show that the PDF-based method is with higher ACC than the PLS-based method in most of the dimensions using linear and SVM-RBF classifiers. For example, for 20 components, the PDF-based ACC performance is 88.50% while PLS ACC performance is 81.96% using SVM-linear. There are few cases in which PLS ACC is higher. The same observation is valid for AUC and SPE, where the PDF-based method is mostly superior to the PLS-based method. On the other hand, although for SEN the PDF-based method is better than the PLS-based method in SVM-linear, the PLS-based method is higher for the SVM-RBF classifier.

4.5. Performance of PDF technique using optimal number of bins

As proposed in Section 3.2.3, the optimal number of bins is determined by maximizing the Fisher criterion applied to the two classes (AD and HC) of the training data in each fold through the cross-validation process. Table 3 presents the average of the performances of the classifiers with the optimal number of bins obtained in each fold, through 10-fold cross-validation. The proposed method of determining the optimal number of components (i.e. the number of bins) is also applied to PLS. By examining the results of Table 3, it was observed that the overall performance of the proposed PDF-based method with the optimal number of

bins is superior to PLS for SVM-linear, where the results of both methods are comparable for SVM-RBF.

4.6. Performance comparison to other methods

Recently, several studies have reported classification results to distinguish AD and HC based on MRI. Zhang et al. [38] used multimodal classification of AD based on the combination of MRI, CSF, and PET. They reported ACCs of 86.2%, 82.1%, and 86.5% in the classification of AD/HC by MRI, CSF, and PET imaging modalities, respectively. Also, they achieved a high accuracy performance (93.2%) by combining the MRI, CSF, and PET results. Querbes et al. [39] achieved an ACC of 85% based on the cortical thickness feature from MRI data. Hinriches et al. [40] reported an ACC of 75.27% based on MRI data and increased it to 81% by combining MRI and PET. Vemuri et al. [41] announced an SEN/SPE of 86/86% in 380 subjects using the STructural Abnormality iNdex (STAND) score from MRI data. Westman et al. [42] presented an ACC of 87% from MRI data and increased it to 91.8% by combining MRI data with CSF measures. Papakostas et al. [43] applied two methods to analyze MRI data, namely, VBM and deformation-based morphometry (DBM), on 98 female subjects. They extracted features based on three different models: MSD, displacement magnitude (DM), and Jacobian determinant (JD). They also investigated their methods with several classifiers. They reported ACCs of 85%, 84%, and 79% for the three models, respectively. Aguilar et al. [44] used FreeSurfer software to compute cortical thickness and volumetric measures, yielding an ACC of 84.9% for the artificial neural network (ANN) classifier from MRI data and of 88.8% for the SVM classifier by combining MRI data with educational and demographic data. Zhou et al. [45] employed FreeSurfer software to calculate 55 volumetric variables from MRI data. They reported an ACC of 78% for MRI data and 92.4% by combining MRI data with the MMSE. Savio et al. [9] studied the feature-extraction process with VBM analysis on 98 female subjects only and achieved the best results with 86% accuracy for the RBF-AB-SVM classifier. Khedher et al. [30] reported an ACC of 88.49% by combining GM and WM modalities in MRI. Klöppel et al. [13] employed

Table 1

Performance comparison on VBM features data sets on 10 fold cross validation for raw feature vectors.

Classifier	ACC (%)	SEN (%)	SPE (%)	AUC (%)
SVM-linear	83.58	82.04	85.12	92.10
SVM-RBF	86.02	89.70	82.35	93.13

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area under curve; SVM, Support vector machine; RBF, Radial basis function.

Table 2

Performance analysis of the PDF based method in comparison to PLS based method.

(a) Performance comparison on PLS reduced features data sets on 10 fold cross validation						(b) Performance comparison on PDF reduced features data sets on 10 fold cross validation						
No. of components	ACC (%)	SEN (%)	SPE (%)	AUC (%)	Classifier	No. of components	ACC (%)	SEN (%)	SPE (%)	AUC (%)	Classifier	
2	87.34	84.65	90.03	95.33	SVM	2	86.19	83.88	88.50	94.85	SVM	
10	85.42	81.57	89.26	93.31		10	87.73	86.19	89.26	95.62		
20	81.96	81.57	82.34	92.25		20	88.50	86.19	90.80	95.50		
30	81.19	80.03	82.34	91.66		30	88.50	88.50	88.50	94.73		
40	81.96	80.03	83.88	92.19		40	86.96	83.88	90.03	94.91		
50	82.73	80.03	85.42	92.49		50	87.34	86.96	87.73	94.91		Linear Kernel
60	82.73	80.03	85.42	92.66		60	88.11	86.96	89.26	95.15		
70	83.88	82.34	85.42	92.90		70	87.34	87.73	86.96	95.21		
80	84.26	82.34	86.19	93.14		80	88.50	89.26	87.73	95.80		
90	85.03	83.88	86.19	93.26		90	86.96	86.96	86.96	94.62		
100	85.03	83.88	86.19	93.31	100	88.50	87.73	89.26	96.21			
2	86.53	88.46	84.61	91.60	SVM	2	88.07	87.69	88.46	95.86	SVM	
10	74.61	96.15	53.07	90.41		10	89.61	89.23	90.00	96.39		
20	79.23	94.61	63.84	93.20		20	88.84	86.92	90.76	96.51		
30	86.76	93.07	78.46	94.50		30	90.00	90.00	90.00	96.09		
40	88.84	92.30	85.38	94.73		40	88.84	88.46	89.23	95.92		
50	88.07	90.76	85.38	95.27		50	89.61	88.46	90.76	96.15		
60	88.46	90.76	86.15	95.38		60	89.61	90.00	89.23	96.27		RBF Kernel
70	90.38	90.76	90.00	95.74		70	88.84	86.92	90.76	96.21		
80	90.76	90.76	90.76	95.86		80	90.00	90.00	90.00	96.75		
90	90.76	90.76	90.76	95.92		90	90.00	88.46	91.53	96.75		
100	90.76	90.76	90.76	95.92	100	90.76	90	91.53	97.04			

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area under curve; SVM, Support vector machine; RBF, Radial basis function.

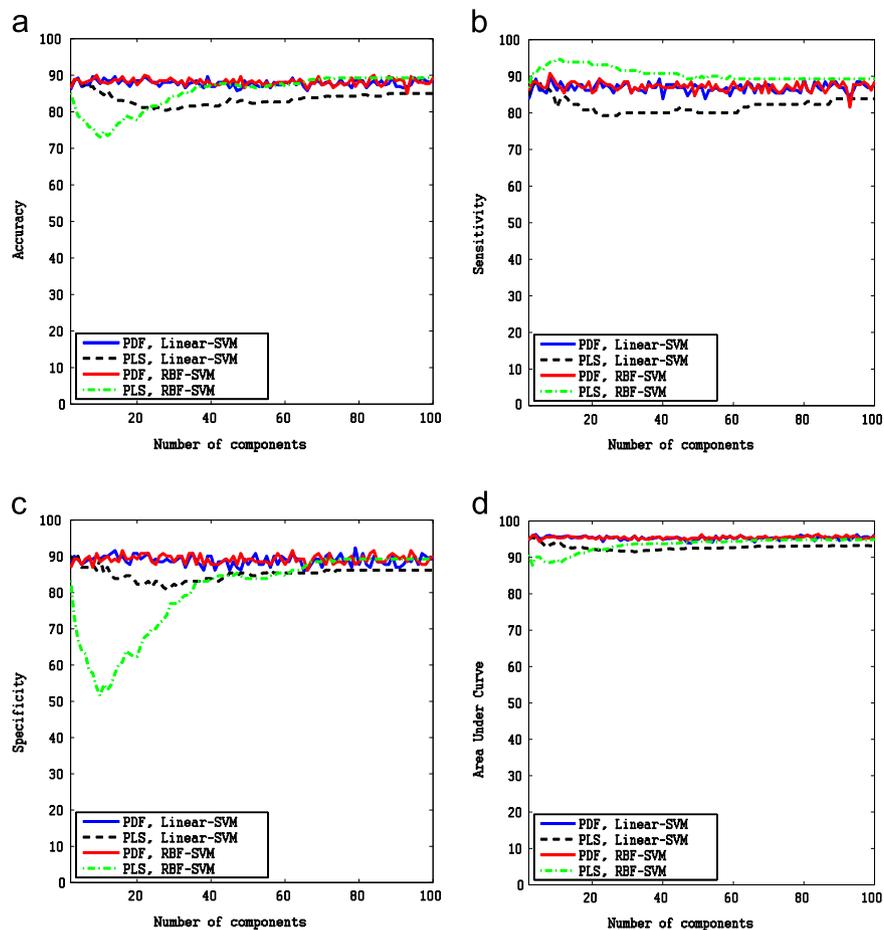


Fig. 6. Classifier performance based on PLS and PDF feature selection: (a) Accuracy, (b) sensitivity, (c) specificity and (d) area under curve.

Table 3

Performance results of the PDF and PLS based methods with optimal number of bins.

Classifier	ACC (%)	SEN (%)	SPE (%)	AUC (%)
PDF-SVM-linear	89.65	87.73	91.57	95.33
PDF-SVM-RBF	88.83	87.73	90.03	95.39
PLS-SVM-linear	85.42	84.65	86.19	93.32
PLS-SVM-RBF	89.26	89.26	89.26	95.09

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area under curve; PDF, Probability distribution function; PLS, Partial least squares; SVM, Support vector machine, RBF, Radial basis function.

leave-one-out as a validation method in three different groups (Groups I, II, and III) with different severity of atrophy in AD. The ACC of Group I was 95%, of Group II was 92.9%, and of Group III was 81.1%. The severity of atrophy in Group I was the highest, making this group the most successful among the three. A study by Cuingnet et al. [46] comprised 10 methods using the ADNI database. They reported a SEN of 81% and a SPE of 95% as the best performances. In this paper, a set of a total of 260 MRI samples was used in the AD and HC groups, with superior results with respect to ACC, SEN, and AUC in Table 4 except for the results of Klöppel et al. [13] for Groups I and II. One of the main reasons for this observation stems from the fact that the severity of the atrophy of Groups I and II was higher than that of Group III and our dataset. Additionally, using the leave-one-out method already gives an advantage to the method employed by Klöppel et al. [13] against the 10-fold cross-validation technique used in the proposed method. The experimental results using the proposed

PDF-based approach with SVM by linear Kernel generates 89.65% accuracy, 87.73% sensitivity, 91.57% specificity, and 95.33% AUC. The details of the parameters used in classification performance with different methods by using MRI data are provided in Table 4. Some of the results reported in Table 4 use ADNI data-set, where the others use different or private data-sets. Additionally, the results from ADNI data-set are using different number of AD/HC samples. In order to have comparable results, we have used ADNI data-set with high number of samples (130 AD and 130 HC), which we believe provides a suitable ground for acceptable comparisons.

5. Conclusion

In this paper, an automatic CAD technique was introduced based on a novel statistical feature-selection process, namely, PDF of VOI, for the classification of AD. The proposed feature-selection method compresses the statistical information of high-dimensional data into a lower-dimensional vector. This approach was used for high-dimensional classification, especially for feature-extracted VOI of gray matter atrophy. The PDF-based feature-selection approach was compared to the standard PLS-based classification using SVM classifiers. The results clearly indicated that the PDF-based feature-selection method is a reliable alternative to the PLS-based method, in which the performance of the proposed PDF-based method with the optimal number of bins is superior to PLS for SVM-linear, and the results of both methods are comparable for SVM-RBF. Moreover, PDF generation does not require complex matrix operations, making the feature-extraction process computationally cheaper than alternative dimensionality-reduction methods, such as PLS. The proposed

Table 4

Supervised classification results of Alzheimer's disease and healthy control subjects on MRI data.

Author	Imaging modality	Source of data	AD/HC	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Zhang et al. [38]	MRI	ADNI	51/52	10 Fold	86.2	86.0	86.3	–
Querbes et al. [39]	MRI	ADNI	130/130	10 Fold	85	–	–	–
Hinrichs et al. [40]	MRI	ADNI	77/82	10 Fold	75.27	63.06	81.86	82.48
Vemuri et al. [41]	MRI	ADRC/ADPR	190/190	4 Fold	–	86	86	–
Westman et al. [42]	MRI	ADNI	96/111	10 Fold	87	83.3	90.1	93
Papakostas et al. [43]	MRI	OASIS	49/49	10 Fold	85	78	92	–
Aguilar et al. [44]	MRI	ADNI	116/110	10 Fold	84.9	80.2	90	88
Zhou et al. [45]	MRI	ADNI	59/127	2 Fold	78.2	68.5	81.7	–
Savio et al. [9]	MRI	OASIS	49/49	10 Fold	86	80	92	–
Khedher et al. [30]	MRI	ADNI	188/229	10 Fold	88.49	85.11	91.27	–
Klöppel et al. [13]	MRI (Group I)	Private	20/20	Leave-one-out	95.0	95.0	95.0	–
Klöppel et al. [13]	MRI (Group II)	Private	14/14	Leave-one-out	92.9	100	85.7	–
Klöppel et al. [13]	MRI (Group III)	Private	33/57	Leave-one-out	81.1	60.6	93.0	–
Cuingnet et al. [46] ^a	MRI	ADNI	162/137	2 Fold	–	81.0	95.0	–
Proposed method	MRI	ADNI	130/130	10 Fold	89.65	87.73	91.57	95.33

^a This paper by Cuingnet et al. compares ten methods and the best performance is given here.

PDF-based method not only extracts the selected statistical features but also reduces the dimensionality of the input vectors to feature vectors with acceptably low dimensions. It is apparent that the computational cost of PDF calculation is negligibly low when compared to PLS. As part of future prospects on PDF-based pattern recognition in neuroimaging, it is suggested to use data fusion techniques for the proposed MRI modality with other modalities, such as PET, CSF, and WM, and to combine them using the proposed PDF-based approach in order to achieve higher accuracy. The PDF-based data fusion technique has already been used successfully in recent studies for the improvement of face-recognition performance [16,17].

Conflict of interest statement

None.

References

- [1] C.L. Carter, E.M. Resnick, M. Mallampalli, A. Kalbarczyk, Sex and gender differences in Alzheimer's disease: recommendations for future research, *J. Women's Health* 21 (2012) 1018–1023. <http://dx.doi.org/10.1089/jwh.2012.3789>.
- [2] F.L. Seixas, B. Zadrozny, J. Laks, A. Conci, D.C. Muchaluat Saade, A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment, *Comput. Biol. Med.* 51 (2014) 140–158. <http://dx.doi.org/10.1016/j.combiomed.2014.04.010>.
- [3] H. Matsuda, S. Mizumura, K. Nemoto, F. Yamashita, E. Imabayashi, N. Sato, et al., Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated Lie algebra improves the diagnosis of probable Alzheimer disease, *Am. J. Neuroradiol.* 33 (2012) 1109–1114. <http://dx.doi.org/10.3174/ajnr.A2935>.
- [4] T. Nakatsuka, E. Imabayashi, H. Matsuda, R. Sakakibara, T. Inaoka, H. Terada, Discrimination of dementia with Lewy bodies from Alzheimer's disease using voxel-based morphometry of white matter by statistical parametric mapping 8 plus diffeomorphic anatomic registration through exponentiated Lie algebra, *Neuroradiology* 55 (2013) 559–566. <http://dx.doi.org/10.1007/s00234-013-1138-9>.
- [5] M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, et al., Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease, *Neurocomputing* 74 (2011) 1260–1271. <http://dx.doi.org/10.1016/j.neucom.2010.06.025>.
- [6] P. Vemuri, C.R. Jack, Role of structural MRI in Alzheimer's disease, *Alzheimers Res. Ther.* 2 (2010) 23. <http://dx.doi.org/10.1186/alzrt47>.
- [7] C.H.C. Huang, B.Y. Bin Yan, H.J.H. Jiang, D.W.D. Wang, Combining voxel-based morphometry with artificial neural network theory in the application research of diagnosing Alzheimer's disease, in: *Int. Conf. Biomed. Eng. Informatics*, 1, 2008, pp. 250–254. <http://dx.doi.org/10.1109/BMEI.2008.245>.
- [8] J.Z.J. Zhang, B.Y. Bin Yan, X.H.X. Huang, P.Y.P. Yang, C.H.C. Huang, The diagnosis of Alzheimer's disease based on voxel-based morphometry and support vector machine, in: *Fourth Int. Conf. Nat. Comput.* 2, 2008, pp. 197–201. <http://dx.doi.org/10.1109/ICNC.2008.804>.
- [9] A. Savio, M.T. García-Sebastián, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga, et al., Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI, *Comput. Biol. Med.* 41 (2011) 600–610. <http://dx.doi.org/10.1016/j.combiomed.2011.05.010>.
- [10] C.C.M. Yanxi Liu, Discriminative MR image feature analysis for automatic Schizophrenia and Alzheimer's disease classification, *Med. Image Comput. Comput. Interv.—MICCAI 2004* (2004) 393–401.
- [11] Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick, C. Davatzikos, Morphological classification of brains via high-dimensional shape transformations and machine learning methods, *NeuroImage* 21 (2004) 46–57. <http://dx.doi.org/10.1016/j.neuroimage.2003.09.027>.
- [12] G. Fung, J. Stoeckel, SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information, *Knowl. Inf. Syst.* 11 (2007) 243–258. <http://dx.doi.org/10.1007/s10115-006-0043-5>.
- [13] S. Klöppel, C.M. Stonnington, C. Chu, B. Draganski, R.I. Scahill, J.D. Rohrer, et al., Automatic classification of MR scans in Alzheimer's disease, *Brain* 131 (2008) 681–689. <http://dx.doi.org/10.1093/brain/awm319>.
- [14] J. Ashburner, K.J. Friston, Voxel-based morphometry—the methods, *NeuroImage* 11 (2000) 805–821. <http://dx.doi.org/10.1006/nimg.2000.0582>.
- [15] C. Cabral, P.M. Morgado, D. Campos Costa, M. Silveira, Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages, *Comput. Biol. Med.* 58 (2015) 101–109. <http://dx.doi.org/10.1016/j.combiomed.2015.01.003>.
- [16] H. Demirel, G. Anbarjafari, Data fusion boosted face recognition based on probability distribution functions in different colour channels, *EURASIP J. Adv. Signal Process.* (2009). <http://dx.doi.org/10.1155/2009/482585>.
- [17] H. Demirel, G. Anbarjafari, Improved PDF Based Face Recognition Using Data Fusion, *J. Electron. Mech. Mechatronics Eng.* 2 (n.d.) 195–200. (http://www.aydin.edu.tr/ijemme/articles/vol2num2/improved_pdf_based_face_recognition_using_data_fusion.pdf).
- [18] H. Demirel, G. Anbarjafari, Pose invariant face recognition using probability distribution functions in different color channels, *Signal Process. Lett.* 15 (2008) 537–540. <http://dx.doi.org/10.1109/LSP.2008.926729>.
- [19] Y. Hirata, H. Matsuda, K. Nemoto, T. Ohnishi, K. Hirao, F. Yamashita, et al., Voxel-based morphometry to discriminate early Alzheimer's disease from controls, *Neurosci. Lett.* 382 (2005) 269–274. <http://dx.doi.org/10.1016/j.neulet.2005.03.038>.
- [20] G.B. Karas, E.J. Burton, S. a R.B. Rombouts, R. a. Van Schijndel, J.T. O'Brien, P. Scheltens, et al., A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry, *NeuroImage* 18 (2003) 895–907. [http://dx.doi.org/10.1016/S1053-8119\(03\)00041-7](http://dx.doi.org/10.1016/S1053-8119(03)00041-7).
- [21] G. Chételat, B. Landeau, F. Eustache, F. Mézenge, F. Viader, V. De La Sayette, et al., Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study, *NeuroImage* 27 (2005) 934–946. <http://dx.doi.org/10.1016/j.neuroimage.2005.05.015>.
- [22] J.C. Baron, G. Chételat, B. Desgranges, G. Percey, B. Landeau, V. de la Sayette, et al., In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, *NeuroImage* 14 (2001) 298–309. <http://dx.doi.org/10.1006/nimg.2001.0848>.
- [23] S. Modi, M. Bhattacharya, N. Singh, R.P. Tripathi, S. Khushu, Effect of visual experience on structural organization of the human brain: a voxel based morphometric study using DARTEL, *Eur. J. Radiol.* 81 (2012) 2811–2819. <http://dx.doi.org/10.1016/j.ejrad.2011.10.022>.
- [24] J. Ashburner, A fast diffeomorphic image registration algorithm, *NeuroImage* 38 (2007) 95–113. <http://dx.doi.org/10.1016/j.neuroimage.2007.07.007>.
- [25] A. Klein, J. Andersson, B. a. Ardekani, J. Ashburner, B. Avants, M.C. Chiang, et al., Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration, *NeuroImage* 46 (2009) 786–802. <http://dx.doi.org/10.1016/j.neuroimage.2008.12.037>.
- [26] J. Cousijn, R.W. Wiers, K.R. Ridderinkhof, W. Van den Brink, D.J. Veltman, A. E. Goudriaan, Grey matter alterations associated with cannabis use: results of

- a VBM study in heavy cannabis users and healthy controls, *NeuroImage* 59 (2012) 3845–3851. <http://dx.doi.org/10.1016/j.neuroimage.2011.09.046>.
- [27] F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, Early diagnosis of Alzheimer's disease based on partial least squares and support vector machine, *Expert Syst. Appl.* 40 (2013) 677–683. <http://dx.doi.org/10.1016/j.eswa.2012.07.071>.
- [28] J. Ramírez, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, et al., Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification, *Neurosci. Lett.* 472 (2010) 99–103. <http://dx.doi.org/10.1016/j.neulet.2010.01.056>.
- [29] R. Chaves, J. Ramírez, J.M. Górriz, C.G. Puntonet, Association rule-based feature selection method for Alzheimer's disease diagnosis, *Expert Syst. Appl.* 39 (2012) 11766–11774. <http://dx.doi.org/10.1016/j.eswa.2012.04.075>.
- [30] L. Khedher, J. Ramírez, J.M. Górriz, A. Brahim, F. Segovia, Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images, *Neurocomputing* 151 (2015) 139–150. <http://dx.doi.org/10.1016/j.neucom.2014.09.072>.
- [31] L. Tang, S. Peng, Y. Bi, P. Shan, X. Hu, A new method combining LDA and PLS for dimension reduction, *PLoS One* (2014). <http://dx.doi.org/10.1371/journal.pone.0096944>.
- [32] Q. Gao, J. Liu, H. Zhang, J. Hou, X. Yang, Enhanced fisher discriminant criterion for image recognition, *Pattern Recognit.* 45 (2012) 3717–3724. <http://dx.doi.org/10.1016/j.patcog.2012.03.024>.
- [33] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, 1988.
- [34] Y. Chen, J. Storrs, L. Tan, L.J. Mazlack, J.H. Lee, L.J. Lu, Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach, *J. Neurosci. Methods* 221 (2014) 22–31. <http://dx.doi.org/10.1016/j.jneumeth.2013.09.001>.
- [35] J. Dukart, K. Mueller, H. Barthel, A. Villringer, O. Sabri, M.L. Schroeter, Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI, *Psychiatry Res.—Neuroimaging* 212 (2013) 230–236. <http://dx.doi.org/10.1016/j.psychres.2012.04.007>.
- [36] X. Song, N.K. Chen, A SVM-based quantitative fMRI method for resting-state functional network detection, *Magn. Reson. Imaging* 32 (2014) 819–831. <http://dx.doi.org/10.1016/j.mri.2014.04.004>.
- [37] W. Lee, B. Park, K. Han, Classification of diffusion tensor images for the early detection of Alzheimer's disease, *Comput. Biol. Med.* 43 (2013) 1313–1320. <http://dx.doi.org/10.1016/j.combiomed.2013.07.004>.
- [38] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment, *NeuroImage* 55 (2011) 856–867. <http://dx.doi.org/10.1016/j.neuroimage.2011.01.008>.
- [39] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Démonet, V. Duret, et al., Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve, *Brain* 132 (2009) 2036–2047. <http://dx.doi.org/10.1093/brain/awp105>.
- [40] C. Hinrichs, V. Singh, G. Xu, S. Johnson, MKL for robust multi-modality AD classification (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), *Lect. Notes Comput. Sci.* (2009) 786–794. http://dx.doi.org/10.1007/978-3-642-04271-3_95.
- [41] P. Vemuri, J.L. Gunter, M.L. Senjem, J.L. Whitwell, K. Kantarci, D.S. Knopman, et al., Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies, *NeuroImage* 39 (2008) 1186–1197. <http://dx.doi.org/10.1016/j.neuroimage.2007.09.073>.
- [42] E. Westman, J.S. Muehlboeck, A. Simmons, Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion, *NeuroImage* 62 (2012) 229–238. <http://dx.doi.org/10.1016/j.neuroimage.2012.04.056>.
- [43] G. a. Papakostas, a. Savio, M. Graña, V.G. Kaburlasos, A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data, *Neurocomputing* 150 (2015) 37–42. <http://dx.doi.org/10.1016/j.neucom.2014.02.076>.
- [44] C. Aguilar, E. Westman, J.S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, et al., Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment, *Psychiatry Res.—Neuroimaging* 212 (2013) 89–98. <http://dx.doi.org/10.1016/j.psychres.2012.11.005>.
- [45] Q. Zhou, M. Goryawala, M. Cabrerizo, J. Wang, W. Barker, D. a Loewenstein, et al., An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment, *IEEE Trans. Biomed. Eng.* 61 (2014) 2245–2253. <http://dx.doi.org/10.1109/TBME.2014.2310709>.
- [46] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.O. Habert, et al., Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database, *NeuroImage* 56 (2011) 766–781. <http://dx.doi.org/10.1016/j.neuroimage.2010.06.013>.