Computers in Biology and Medicine 83 (2017) 109–119

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm

Iman Beheshti[a,*], Hasan Demirel[b], Hiroshi Matsuda[a], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, 4-1-1, Ogawahigashi-cho, Kodaira, Tokyo 187-8551, Japan
[b] Biomedical Image Processing Group, Department of Electrical & Electronic Engineering, Eastern Mediterranean University, Famagusta, Mersin 10, Turkey

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment conversion
Feature ranking
Genetic algorithm

ABSTRACT

We developed a novel computer-aided diagnosis (CAD) system that uses feature-ranking and a genetic algorithm to analyze structural magnetic resonance imaging data; using this system, we can predict conversion of mild cognitive impairment (MCI)-to-Alzheimer's disease (AD) at between one and three years before clinical diagnosis. The CAD system was developed in four stages. First, we used a voxel-based morphometry technique to investigate global and local gray matter (GM) atrophy in an AD group compared with healthy controls (HCs). Regions with significant GM volume reduction were segmented as volumes of interest (VOIs). Second, these VOIs were used to extract voxel values from the respective atrophy regions in AD, HC, stable MCI (sMCI) and progressive MCI (pMCI) patient groups. The voxel values were then extracted into a feature vector. Third, at the feature-selection stage, all features were ranked according to their respective *t*-test scores and a genetic algorithm designed to find the optimal feature subset. The Fisher criterion was used as part of the objective function in the genetic algorithm. Finally, the classification was carried out using a support vector machine (SVM) with 10-fold cross validation. We evaluated the proposed automatic CAD system by applying it to baseline values from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (160 AD, 162 HC, 65 sMCI and 71 pMCI subjects). The experimental results indicated that the proposed system is capable of distinguishing between sMCI and pMCI patients, and would be appropriate for practical use in a clinical setting.

1. Introduction

Alzheimer's disease (AD) is a form of progressive irreversible dementia that occurs most frequently in older adults. AD gradually defaces the regions of the brain that are responsible for memory, thinking, learning, and other cognitive abilities [1]. It has been estimated that the number of patients who suffer from AD will double in the next two decades, and will reach 13.8 million people by 2050 [2]. In America, AD is one of the top 10 causes of death that cannot be cured or prevented [1]. Early detection may help to clarify the mechanisms underlying AD and to improve the quality of life for AD patients [1,3].

Mild cognitive impairment (MCI), a prodromal stage of AD, is a high-risk dementia condition in which acquired cognitive deficiency has no significant effect on the functional activities of daily living [2]. Patients with MCI have an increased risk of eventually developing AD. Recent studies have shown that up 20% of people aged 65 and older may have MCI [2]. Recently, neuroimaging techniques, such as structural magnetic resonance imaging (sMRI) [4–11], functional MRI [12–14], diffusion tensor imaging [15–17], positron emission tomography (PET), and single photon emission computed tomography (SPECT) [18–21] have been used successfully for AD classification and the prediction of MCI-to-AD conversion. Despite recent developments in the early detection of AD, the prediction of disease progression is still challenging and requires further investigation.

In this study, we present a computer-aided diagnosis (CAD) system

* Corresponding author.
E-mail address: Iman.beheshti@ncnp.go.jp (I. Beheshti).

[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.

http://dx.doi.org/10.1016/j.compbiomed.2017.02.011
Received 1 November 2016; Received in revised form 26 January 2017; Accepted 23 February 2017

that applies voxel-based morphometry (VBM) analysis to sMRI data in order to make an early prediction of conversion to AD at between one and three years before clinical diagnosis. Such prediction is useful in determining a subject's clinical treatment plan or in judgment of the clinical trial for basic drugs to treat AD. Data from sMRI of the brain have been widely used to predict early conversion to AD, and VBM analysis is a widely used approach for this purpose [11]. Many researchers have used sMRI feature extraction for AD classification or the prediction of MCI-to-AD conversion using a number of different methods that include morphometry [22–24], measurement of the hippocampus and the medial temporal lobe [25–31], regions of interest (ROI)/volume of interest (VOI) [32–34], and gray matter (GM) voxels in the automatic segmentation of images [35]. In this paper, we propose an automatic feature-selection method based on feature ranking and a genetic algorithm (GA) for AD classification and prediction of MCI-to-AD conversion, which are the two most difficult tasks in AD detection studies. The proposed feature-selection method was realized by extracting voxel-values as raw-feature data from VOIs obtained from VBM analysis. The extracted raw-feature vectors were then reduced to lower-dimensional feature vectors using the proposed feature-selection method.

In addition, we extracted features from stable MCI (sMCI) and progressive MCI (pMCI) samples based on GM atrophy patterns in AD subjects and HCs. We then evaluated the performance of the proposed method on baseline data from 458 samples from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (160 AD, 162 HC, 65 sMCI and 71 pMCI subjects). In a direct comparison, the proposed CAD system exhibited better performance than classifiers using all raw features and principal component analysis (PCA) data reduction. The accuracy of the proposed CAD system for classifying AD/HC and sMCI/pMCI using baseline sMRI data was 93.01% and 75%, respectively. In addition, when compared to state-of-the-art techniques for predicting MCI-to-Alzheimer's conversion, the experimental results show that the performance of the proposed CAD system was comparable to that of alternative methods using only baseline MRI data.

2. Materials

2.1. Image acquisition and samples

The baseline MRI scans and data used in this work were obtained from 3 T T1-weighted images, which were typically 240 × 256 × 176 voxels with a voxel size of 1 mm × 1 mm × 1.2 mm. MR image corrections such as GradWarp and N3 were also applied to correct the geometric distortion caused by gradient non-linearity [36] and bias field, respectively, and to balance the lack of homogeneity due to the absence of uniformity in the radio frequency receiver coils [37]. We divided the dataset into two main groups. The aim was to separate the data used for VBM-based 3D mask generation from the data used for classification. Group 1 (including 68 AD patients and 68 healthy controls (HCs)) was used to model the VBM 3D mask. Group 2 (including 92 AD, 94 HC, 65 stable MCI (sMCI) and 71 progressive MCI (pMCI) subjects) was used for feature extraction, classification and validation. The samples in Groups 1 and 2 were totally independent. We used only baseline data.

2.2. Defining volumes of interest (VOIs) (Group 1)

Group 1 consisted of 136 subjects divided into HC and AD cohorts. Table 1 shows details of the demographics and clinical characteristics of the samples used in Group 1. We used this group in the VBM analysis to find regions of GM atrophy in patients with AD compared to the HCs. These regions were segmented as a 3D mask to extract voxel values in the feature-extraction stage. The data used to define the VOI and respective 3D mask were the same as previously reported [38].

Table 1
Characteristics of the subjects used to model the VBM 3D mask (Group 1).

	AD (n=68)	HC (n=68)
Age (mean)	74.33 ± 6.41	74.14 ± 4.95
Range	[60–82]	[66–84]
MMSE (mean)	22.83 ± 2.56	29.38 ± 0.71
Range	[16–25]	[28–30]
CDR (mean)	0.75 ± 0.41	0.0 ± 0
[0/0.5/1/2]	[0/44/19/5]	[68/0/0/0]
f/m	34/34	34/34

Note: AD, Alzheimer's Disease patients; CDR, Clinical Dementia Rating; HC, Healthy Control patients; MMSE, Mini-Mental State Examination; F, Female; M, Male.

2.3. Feature extraction and classification (Group 2)

Group 2 consisted of 322 samples that were divided into 4 cohorts: 1) AD subjects; 2) HC subjects; 3) sMCI subjects, where MCI had been diagnosed for at least 36 months; and 4 pMCI subjects, where there was a baseline MCI diagnosis but conversion to AD had occurred within 1, 2 or 3 years since the baseline. Details of the demographic and clinical characteristics of the samples used in this paper are shown in Table 2.

3. Methodology of the proposed computer-aided design (CAD) system

In this section, we introduce the methodology used to develop the proposed automatic CAD system, based on feature ranking, followed by a GA for early detection of conversion from MCI to AD. First, the 3D T1-weighted MRI data samples from Group 1 were processed using VBM analysis to identify GM regions that were significantly more atrophied in the AD patients than the HCs. The regions of GM atrophy were used to form VOIs, which defined the 3D mask to be used for feature extraction. We then used the 3D mask generated using the Group 1 samples from the first stage of the CAD process, i.e., the derivation of VOIs, to extract voxels as raw features from the subjects in Group 2. In the training set, a statistical *t*-test was then used to rank the raw features. Features with *t*-test values higher than 70% of the highest *t*-test value were considered to be a subset of the top-ranked features of the entire set to be used by the GA to find the optimal features. This threshold provided a good trade-off between the computational cost and accuracy. This subset (i.e., the top-ranked features) was applied to the GA with the Fisher criterion as part of the objective function to identify the optimal distinctive features with the minimum number of selected features. The Fisher criterion helped in selecting an optimal subset of features with the most distinctive information for the classification process. Finally, a linear SVM classifier was used to evaluate the proposed technique. Fig. 1 illustrates

Table 2
Characteristics of the subjects used for feature extraction and classification (Group 2).

	AD (n=92)	HC (n=94)	sMCI (n=65)	pMCI (n=71)
Age (mean)	75.32 ± 6.46	73.35 ± 5.68	70.86 ± 7.04	74.27 ± 7.78
Range	[60–90]	[55–81]	[55–88]	[56–88]
MMSE (mean)	23.45 ± 2.45	29.73 ± 3.85	27.80 ± 1.48	26.05 ± 2.87
Range	[18–25]	[27–30]	[25–30]	[23–30]
CDR (mean)	0.72 ± 0.39	0.00 ± 0.00	0.50 ± 0.00	0.50 ± 0.00
[0/0.5/1/2]	[0/58/30/4]	[94/0/0/0]	[0/65/0/0]	[0/71/0/0]
f/m	50/42	50/44	39/26	44/27

Note: All data are presented in mean ± standard deviation mode. AD, people with Alzheimer's disease; CDR, Clinical Dementia Rating; HC, healthy control participants; sMCI, stable Mild Cognitive Impairment; pMCI, progressive Mild Cognitive Impairment; MMSE, Mini-Mental State Examination; F, Female; M, Male.

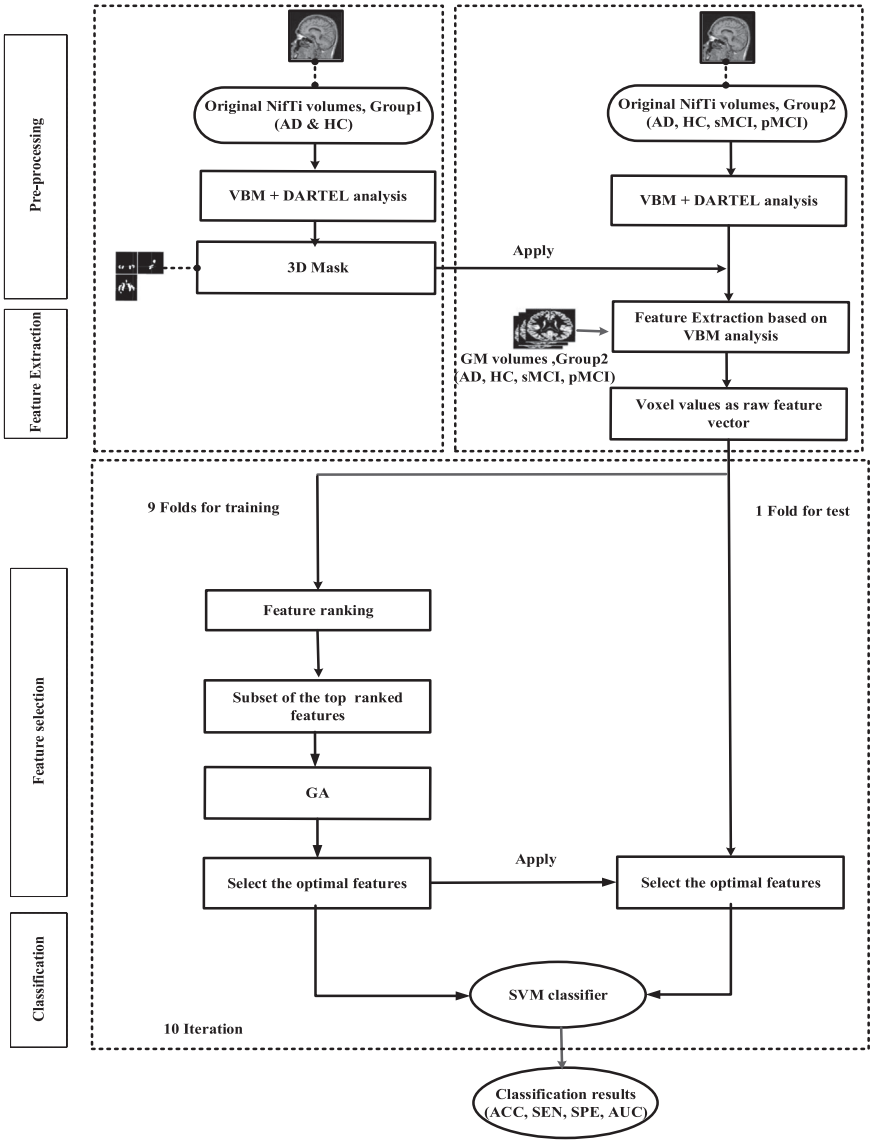

Fig. 1. Pipeline of the proposed automatic CAD system.

the framework of the proposed automatic CAD system.

3.1. Magnet resource imaging (MRI) preprocessing

Preprocessing of the MRI data was carried out using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) and the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm). VBM is an advanced technique that is used to assess the structure of the whole brain with voxel-by-voxel comparisons between subject groups to distinguish degenerative diseases with dementia [39]. Several recent studies have used the VBM method for the early detection of atrophic changes in AD [39–43]. In the present study, the Diffeomorphic Anatomic Registration Through Exponentiated Lie (DARTEL) approach was used as a part of the VBM analysis to enhance inter-subject registration of the MRI images [44,45]. In the VBM8 toolbox, all 3D MR images were corrected for the bias field with regard to homogeneity. The corrected images were normalized and then segmented into GM, white matter (WM), and cerebrospinal fluid (CSF) components. Using a nonlinear deformation, the normalized segmented images were modulated. Only GM components were considered in the current study. All GM components were spatially smoothed with an 8 mm full-width half-maximum Gaussian kernel. After spatial preprocessing, the smoothed, modulated, DARTEL-warped, and normalized GM datasets from Group 1 were used for statistical analysis, whereas datasets from Group 2 were used in the feature extraction phase. The levels of atrophy in regional GM volume were generated by voxel-based analysis over the whole brain for Group 1 subjects. In this manner, the smoothed, modulated, DARTEL-warped, and normalized GM datasets of Group 1 were subjected to the general linear model to detect GM volume changes using voxel-wise two-sample *t*-tests in SPM8. Age was added to the matrix design as a nuisance variable. Absolute threshold masking was adjusted at 0.1 to avoid possible edge effects among GM, WM or CSF components. Significance was set at $p < 0.01$, with correction for family-wise error (FWE), and the threshold was set at 1400 voxels for two-sample comparisons. Fig. 2 illustrates the processing framework of the VBM analysis used to identify significant levels of GM atrophy in the AD relative to the HC samples in Group 1 through SPM8.

3.2. Feature extraction

The feature-extraction procedure based on the preprocessing analysis was then applied to isolate the VOIs. Brain regions with significantly decreased GM volumes, obtained using VBM analysis in

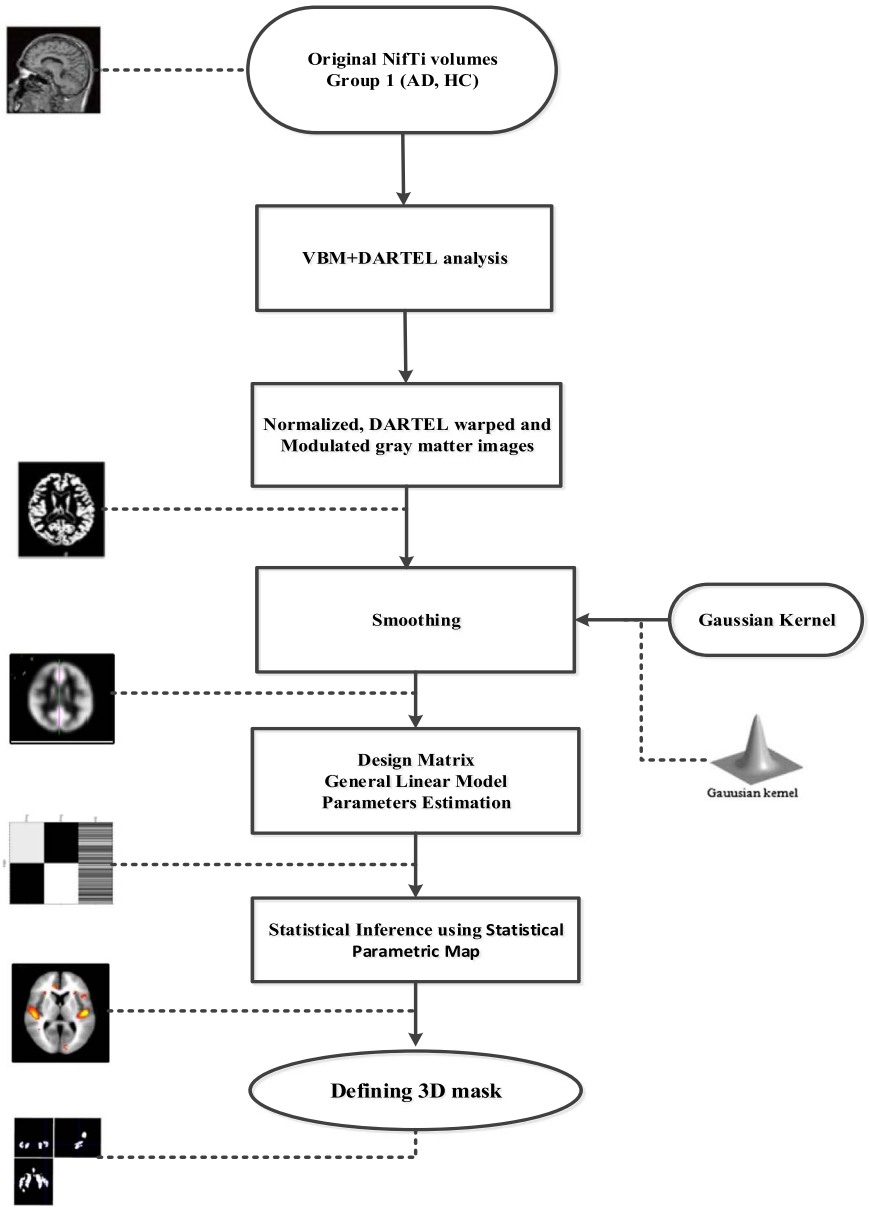

Fig. 2. The VBM plus DARTEL processing pipeline to reveal significant gray matter atrophies among group 1 samples using SPM8.

AD patients relative to HCs in Group 1 subjects, were segmented as a 3D mask. Because sMCI samples appear more healthy, whereas pMCI samples are closer to AD [46], we decided to use the 3D mask not only to extract features for AD and HC samples but also for pMCI and sMCI subjects. The mask was therefore applied to all the smoothed, normalized, modulated, DARTEL-warped, GM density volumes obtained from the preprocessing stage in Group 2 subjects (92 AD, 94 HC, 65 sMCI and 71 pMCI subjects) in order to extract voxel values from atrophy regions as raw-feature vectors.

3.3. Feature selection

As the number of samples available for training is generally very small in comparison to the dimensionality of feature vectors extracted from MRI data (e.g., 36,529 voxels), training an accurate classifier is a challenging process. Many recent studies have explored different feature-selection methods to address this issue. In this study, we chose a novel feature-selection method for high-dimensional data, which combines feature ranking with a GA to reduce the dimensionality, and to select optimal features for the high performance MCI conversion prediction and AD classification. The performance of this method was evaluated using PCA data reduction and raw-feature vectors.

3.3.1. Feature reduction based on principal component analysis (PCA)

The aim of the feature-reduction algorithm was to make available a set of new features to generate a low-dimensional representation of the original data. PCA is a statistical feature-reduction method, which helps extract a set of orthogonal principal components (PCs) from an original dataset [47,48]. Linear combinations of PCs are used to represent high-dimensional original data. Let $\Gamma = [f^1, f^2, ...,f^M]$ be a feature dataset including M features, where $f^j = (x_1^j, x_2^j, ...,x_N^j)^T$ is a vector of values of a feature f^j and N represents the number of samples. PCs are eigenvectors of the gamma matrix of data X as follows:

$$\sum = \frac{1}{N}(\Gamma - \overline{\Gamma})^T(\Gamma - \overline{\Gamma}), \quad (1)$$

where $\overline{\Gamma}$ is the mean vector of each feature. Each PC is measured by its corresponding eigenvalue.

In this study, 10-fold cross validation was used to measure the performance of the classifiers. Ten-fold cross validation takes 90% of data through the training process and 10% through the testing of each clinical group (i.e., the AD/HC and sMCI/pMCI groups). The number of PCs, k, used to generate the projection vectors of the training and testing data was chosen based on the maximum number of training samples for the AD/HC and sMCI/pMCI pairs, respectively.

3.3.2. Feature ranking

Feature ranking helps evaluate the relevance of features and class variables in order to choose the most distinctive features. This method is effective in data with very high dimensionality [49]. Many researchers have investigated different feature-ranking methods as part of feature selection in the pattern recognition field [50–56]. Let $\Gamma = [f^1, f^2, ...,f^M]$ be a feature dataset including M features. The vector $f^j = (x_1^j, x_2^j, ...,x_N^j)^T$ is a vector of values of a feature f^j where x_i^j of this vector shows a feature value of the sample i^{th} and N is the number of samples. Applying a feature ranking algorithm to the feature data-set Γ generates a list of the features $\Omega = [f_*^1, f_*^2, \ldots f_*^M]$ ordered by reduction importance. In this study, we used a statistical indicator, namely the t-test, to measure the level of separation/discrimination between two classes. T-test feature ranking has been used successfully in a number of machine learning studies [57–59]. The t-test measures the statistical significance of the value differences between two classes, as follows:

$$t - value = \frac{\mu_{c1} - \mu_{c2}}{\sqrt{\frac{\sigma_{c1}^2}{n_{c1}} + \frac{\sigma_{c2}^2}{n_{c2}}}}, \quad (2)$$

where μ_{c1}, σ_{c1}^2, n_{c1} and μ_{c2}, σ_{c2}^2, n_{c2} are the mean, variance values, and number of samples of two classes, c_1 and c_2, respectively. To select the top informative features, all features were ranked according to the absolute value of their t-values. Based on feature ranking, we were able to choose the top discriminative q ranked features $[f_*^1, f_*^2, \ldots f_*^q]$, where $q \leq M$.q could be adjusted by the user or determined experimentally. In the present study, q was adjusted with features that were higher than 70% of the maximum t-test values of the training set in each fold. This threshold provided a good trade-off between computational cost and accuracy.

3.3.3. Feature selection using a genetic algorithm (GA)

In this section, we describe the GA used to find the optimum feature subset from the candidate features. We used a binary GA in this study. A chromosome was encoded as an array with length q bits containing binary numbers, as shown in Fig. 3. When *chromosome* $[i^{th}]$ is 1, the i^{th} feature is selected as the optimal feature from the candidate features, otherwise the i^{th} feature is not selected. The main components of the GA comprised population initialization, fitness evaluation, selection, crossover and mutation operations, and termination judgment. The pipeline of the GA is illustrated in Fig. 4. The details of the main components are given below:

a) Initial population. In the binary GA, the population consisted of all the chromosomes in each generation, which were coded as a binary string. In this study, we used a random population base for population initialization.
b) Fitness function. In order to select the optimal feature subset from the candidate features, the following objective function was proposed:

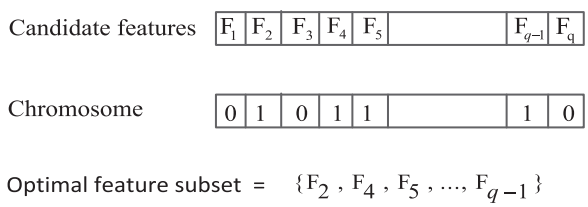

Fig. 3. Feature representation as *chromosome* in Binary GA.

$$\min Z$$
$$Z = E(1 + \alpha . RF)' \quad (3)$$

where E is based on the Fisher criterion and calculated in the training part of each fold, α is a constant coefficient and RF is the ratio of the number of selected features (number of selected features/length q). The objective function finds the optimal features with the maximum discriminative and minimum number of features. In this study, the value of α was adjusted to 0.5. The Fisher criterion evaluates the class separation between two groups of data and also helps to find the optimal features with the most discriminative information for the classification process. The Fisher criterion was calculated as follows:

$$FC = \frac{trace(S_B)}{trace(S_W)}, \quad (4)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix [60]. For two classes, c_1 and c_2, the between-class scatter and within-class scatter matrices are defined as [62,63]:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (5)$$

$$S_W = \sum_{X_i \in c_1} (X_i - \mu_1)(X_i - \mu_1)^T + \sum_{X_i \in c_2} (X_i - \mu_2)(X_i - \mu_2)^T \quad (6)$$

where μ_1 is the mean of the data in class 1 and μ_2 is the mean of the data in class 2. Because the proposed objective function is based on minimization, we used $FC^{-1} = \frac{trace(S_W)}{trace(S_B)}$ as E in Eq. (3).

c) Selection operator. The aim of selection operation is to select the parent individuals, which will participate in producing offspring for the next generation. In this study, we used a roulette wheel selection technique [50].
d) Crossover operator. The goal of the crossover operator is to generate new individuals by recombining the genes of the chromosomes. In the proposed GA, the following crossover operators were used: single point, double point, and arithmetic. For each iteration in the GA process, one of the above crossover operators was randomly selected based on the wheel selection technique.
e) Mutation operator. The mutation operator is normally initiated after the crossover process by randomly inverting one bit of an individual's chromosome to generate a child.
f) Termination judgment. In the proposed feature-selection GA, the termination algorithm was either based on a predefined number of iterations or based on β iterations, where the difference between two objective functions does not exceed ε. In this study, the values of ε and β were adjusted to 10^{-3} and 30, respectively.

Algorithm 1 outlines the pseudo-code used in the proposed method.

Algorithm 1.. The procedure for determining the optimal subset of features based on the feature ranking and genetic algorithm.

- $V \leftarrow component\_set\ (Data_{Train}, Label_{Train})$
- *Ranked features* ← *t-test feature ranking (V)*
- *Top ranked features* ← *The features higher than 70% of the maximum t-test values*
- $Z \leftarrow \infty$
- *Generate an Initial population*
- *While not terminated condition*
- $(S_w, S_B) \leftarrow compute\_scatter\ (selected\_features, Label_{Train})$
- $E \leftarrow \frac{trace(S_W)}{trace(S_B)}$
- $q \leftarrow$ *length of top-ranked features*
- $RF \leftarrow$ *number of selected features/q*
- $\alpha \leftarrow 0.5$

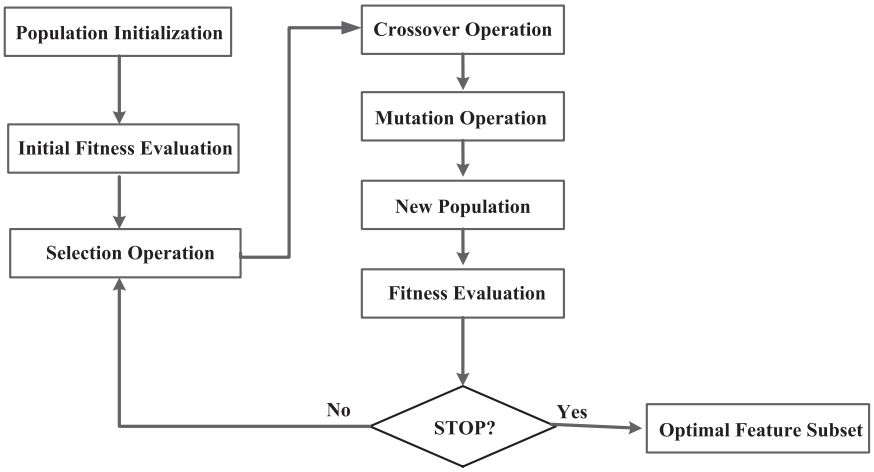

Fig. 4. Flowchart of the Genetic Algorithm.

- $Z \leftarrow E\,(1 + \alpha.\ RF)$
- *Calculate* $Z \leftarrow$ *for each chromosome from the population*
- *Make the next population*
  1. *Selection operator*
  2. *Crossover operator*
  3. *Mutation operator*
- *End while*
- *Find the chromosome with minimum Z*
- *Select the optimal- feature-subset*

3.4. Classifier and performance evaluation

The classifier was operated using an SVM algorithm, a supervised learning method that has been widely used for AD classification problems [6,38,46,62–64]. In this study, an SVM algorithm using LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) with a linear kernel was used for the AD classification and the prediction of MCI-to-AD conversion. In order to achieve a reliable measurement, all performance results were obtained through 10-fold cross validation. The classification results were calculated by means of accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC). These parameters are defined as follows:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$SEN = \frac{TP}{TP + FN} \tag{8}$$

$$SPE = \frac{TN}{TN + FP}, \tag{9}$$

where TP, TN, FN, and FP stand for true positive, true negative, false negative, and false positive, respectively.

4. Experimental results

The experimental data consisted of 458 samples from the ADNI dataset. We first arrived at the results of the preprocessing stage using a VBM and DARTEL analysis on 3D T1-weighted MRI scans to reveal the significance of the difference in volumetric regions with atrophy between the AD subjects in Group 1 and the HCs in the same group. Secondly, we used a 10-fold cross validation process to compare the performance of the proposed feature-selection method for AD classification and the prediction of MCI-to-AD conversion between a trial with PCA data reduction and a trial without feature selection (i.e., using all features). ACC (%), SEN (%), SPE (%), and AUC (%) performance metrics were used to assess the different scenarios.

4.1. Voxel-based morphometry on gray matter (GM) (Group 1)

The VBM and DARTEL analysis revealed significant GM atrophy in the right and left hippocampi, the right inferior parietal lobule, and the right anterior cingulate in patients with AD compared to the HCs in Group 1. Fig. 5 shows the significant regions of GM atrophy in the Group 1 subjects. The voxel locations of these significantly atrophied regions were segmented as a 3D mask. This mask was applied to the GM density volume results from the segmentation step in the VBM-DARTEL analysis of Group 2 to extract a total of 36,529 voxel values as a raw-feature vector.

4.2. Classification results for Alzheimer's disease (AD) and healthy control (HC) subjects

As described in Section 3.3, the proposed feature-selection method was accomplished with a feature-ranking strategy and by using a GA to find optimal top discriminative feature data from raw-feature vectors. The parameters of GA, such as the number of populations, number of iterations, probability of crossover and probability of mutation, were

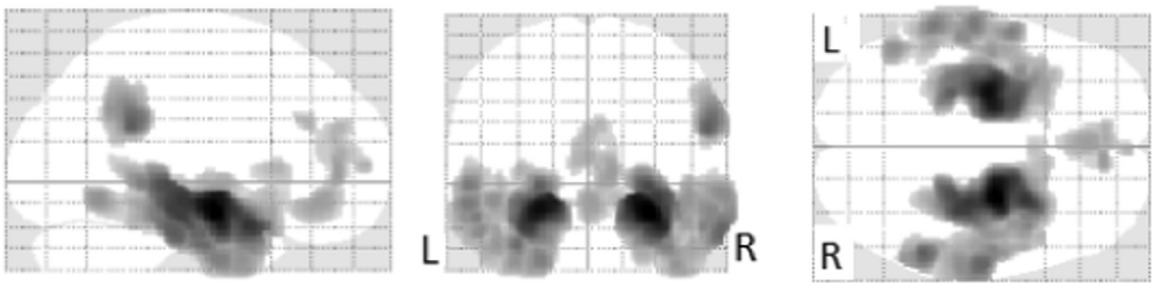

Fig. 5. Significant gray matter atrophy regions in subjects of group 1 [The image is reproduced from a previous study [38]].

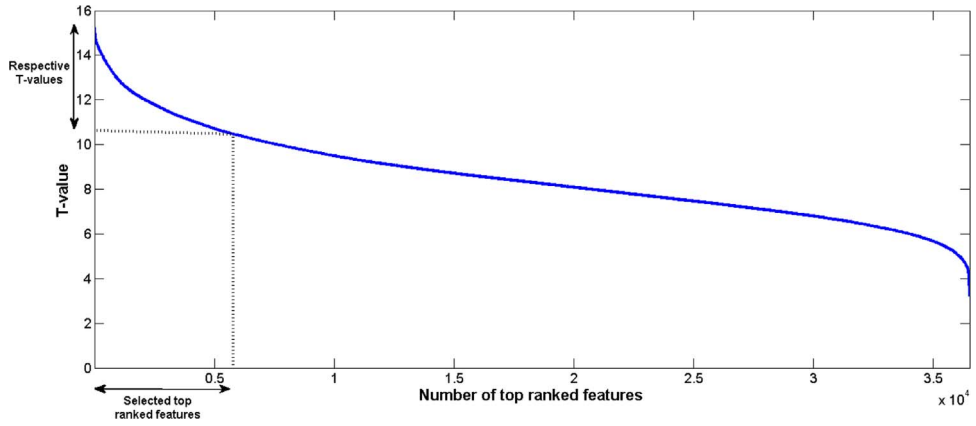

Fig. 6. *t*-test values for the respective ranked features in fold 1 training of AD/HC classification.

Table 3
Classification results in fold 1, HC vs AD.

Feature selection	Number of features	ACC (%)
NO	36529	78.94
Yes (proposed method)	1913	94.73

Note: ACC, Accuracy; AD, people with Alzheimer's disease; HC, healthy control participants.

Table 4
Comparison of the performances of the proposed feature selection method with raw-feature vectors and PCA data reduction for AD classification.

	Feature Selection	ACC (%)	SEN (%)	SPE (%)	AUC (%)
HC vs AD	No	87.63	85.86	89.36	88.70
	PCA	88.70	89.36	88.04	88.70
	Proposed method	93.01	89.13	96.80	93.51

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; AD, people with Alzheimer's disease; HC, healthy Control participants; PCA, Principal Component Analysis.

set at 50, 200, 0.8 and 0.3, respectively. As an example, Fig. 6 shows all *t*-test values for the data in the training set from fold 1 of the HC vs AD samples in Group 2. The maximum *t*-test value is 15.24. Based on the proposed approach to choose the features higher than 70% of the maximum *t*-test values, the features with respective *t*-test values between 10.67 and 15.24 were selected. As a result, 5101 top-ranked features were selected as the top subset of the most discriminative features. This subset was applied to the GA to find the optimum subset based on GA feature selection, with the Fisher criterion as part of the objective function. A total of 1913 features were selected as belonging to the optimal and high-performance feature subset. Table 3 shows the accuracy performance of data from fold 1 of the HC vs AD samples in Group 2 for raw-feature vectors together with the proposed feature selection based on feature ranking and the GA. As can be seen by the data in Table 5, the proposed method is able to improve the performance accuracy to a significant degree, and can also minimize the number of selected features to speed up the learning process. Table 4 presents the average of the performance obtained by 10-fold cross validation between the AD and HC samples from Group 2. With 186 samples for classification in Group 2 (i.e., 94 HCs and 92 AD subjects), a 10-fold cross validation suggested 167 PCs through the PCA process for AD and HC classification. Thus, the number of PCs was chosen as k=167 to generate the projection vectors for the training and testing data. To evaluate the overall performance of the proposed method, PCA data reduction and raw features, a receiver operating characteristic (ROC) curve for AD and HC classification was generated and is shown in Fig. 7(a). Looking at Table 4 and Fig. 7(a), it is clear that the proposed feature-selection method, based on feature ranking and the GA, improves performance to a significant degree when compared to classification using raw features or PCA data reduction. For example, the mean accuracy increases from 87.63% to 93.01% for AD/HC. Our accuracy of 93.01% for AD/HC compares well with other recent classification experiments for separating AD data from HC results, based on MRI data [8,63,65].

4.3. Prediction results for the conversion of stable and progressive MCI to AD

As described in Section 3.2, the GM atrophy pattern between AD subjects and HCs was used to extract significant features from sMCI and pMCI subjects. Table 5 presents the ACC, SEN, SPE, and AUC measurements obtained from 10-fold cross validation for the MCI conversion prediction between sMCI and pMCI subjects. With 136 samples (i.e., 65 sMCI and 71 pMCI subjects), a 10-fold cross validation suggested 122 PCs through the PCA process for this task. Thus, the number of PCs was chosen as k=122 to generate the projection vectors for the training and testing data. Fig. 7(b) shows the ROC curves related to the proposed method, PCA data reduction and raw features for the sMCI and pMCI conversion prediction. As can be seen in Table 5 and Fig. 7(b), the proposed feature-selection method, based on feature ranking and the GA, consistently outperforms the PCA data reduction or raw features for MCI conversion prediction. For example, the proposed method had an ACC performance of 75%, whereas PCA-based data reduction and the raw-feature vector methods had ACCs of 69.11% and 68.38%, respectively, for distinguishing sMCI subjects from pMCI subjects. The 75% accuracy of our proposed method with regard to sMCI and pMCI conversion within 36 months compares well with other recent classification experiments on the separation of sMCI from pMCI results using MRI data [8,10,46,63,65–68].

5. Discussion

Many recent studies have investigated the use of advanced pattern analysis methods to extract complex spatial patterns from the brain structure [13,35,66,69,70]. This is especially the case for high-dimensional pattern analysis in a number of neuroimaging studies [6,33,71,72]. For example, in one study [71], the authors introduced an SVM-recursive feature elimination (SVM-RFE) technique for feature ranking, and used SVM for classification. In another study [6], the authors introduced a statistical feature-selection method based on the probability distribution function (PDF) of the VOI to generate a statistical pattern of the VOI to represent the entire sMRI. In the present study, we introduced an automatic feature-selection method based on feature ranking and a GA for the VOI, which is able to select

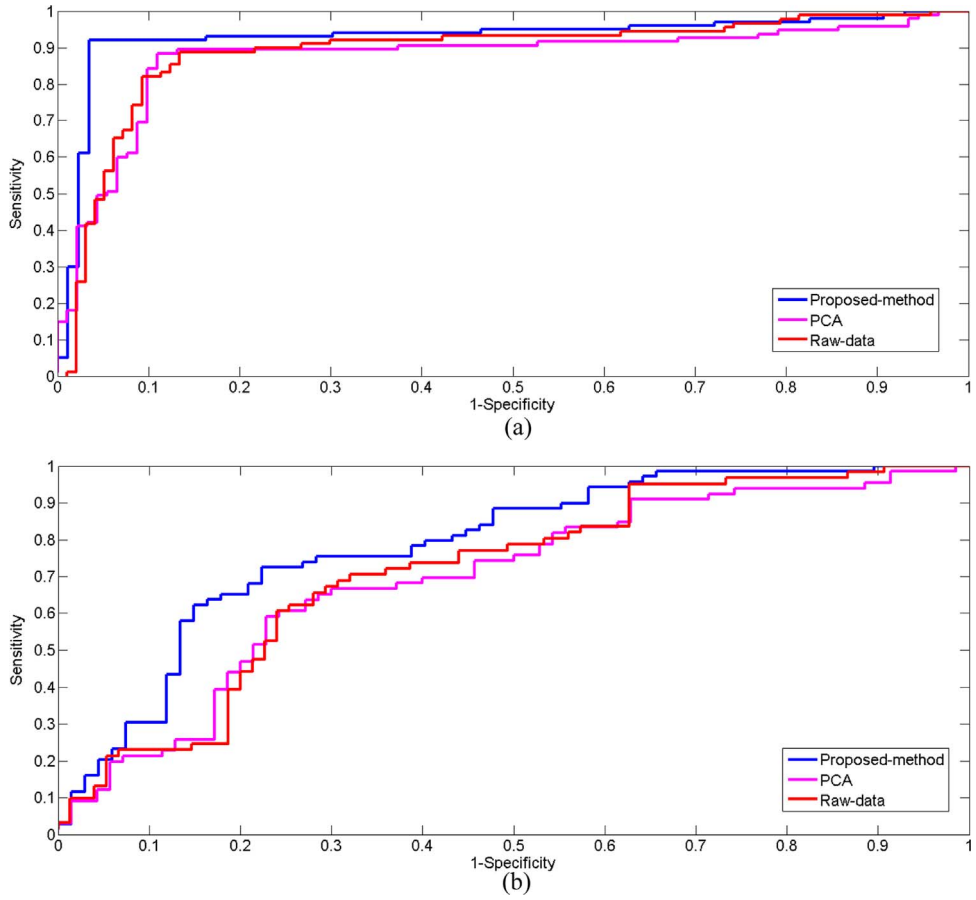

Fig. 7. The ROC curve of the proposed method, PCA data reduction and raw-feature-vectors. (a) AD/HC and (b) pMCI/sMCI.

Table 5
Comparison of the performances of the proposed feature selection method with raw-feature vectors and PCA data reduction for MCI conversion prediction.

	Feature selection	ACC (%)	SEN (%)	SPE (%)	AUC (%)
pMCI vs sMCI	No	68.38	67.69	69.01	68.93
	PCA	69.11	64.61	73.23	68.38
	Proposed method	75.00	76.92	73.23	75.08

Note: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; AUC, Area Under Curve; sMCI, stable Mild Cognitive Impairment; pMCI, progressive Mild Cognitive Impairment; PCA, Principal Component Analysis.

the most discriminative features from a high-dimensional pattern into a lower-dimensional space. In the proposed feature-selection strategy, the Fisher criterion was employed as part of the objective function to find the most discriminative features as an optimal subset. The Fisher criterion helps to select the optimum subset of features with maximum separation between two groups. Using a GA, the proposed feature-selection method not only selects the top discriminative features but also minimizes the dimensionality of the input vectors to low-dimensional space. The experimental results indicate that the proposed feature-selection approach is suitable for high-dimensional pattern recognition, especially for MCI conversion prediction and AD classification. Recent studies have shown that MCI (including both pMCI and sMCI) is a heterogeneous condition where sMCI patients appear more healthy and pMCI patients appear closer to AD [46]. In addition, most of the existing methods for MCI conversion prediction use data taken directly from MCI subjects, ignoring data from other related domains (i.e., AD subjects and HCs) that may provide significant additional data predicting MCI conversion [46]. In this study, we defined a VBM mask based on regions of GM atrophy from AD and HC subjects for feature extraction using pMCI and sMCI subjects. The experimental results demonstrate that extracting features from sMCI and pMCI samples based on GM atrophy patterns in AD subjects and HCs through VBM analysis can also be useful for MCI prediction. Although the proposed method demonstrated a desirable performance for AD detection and MCI conversion prediction, some limitations should be considered in the present study. First, we evaluated the proposed method on an ADNI dataset. In the future, it will be important to evaluate our method using other datasets. Second, we used a small data-sample size, especially for to predict the conversion from MCI to AD (i.e., 65 sMCI and 71 pMCI). This would be worth evaluating on a large dataset in the future. As part of a future study, we suggest employing other meta-heuristic optimization algorithm such as simulated annealing, particle swarm optimization and ant colony optimization for AD classification and MCI conversion prediction. Another priority for future studies will be to use volumetric subcortical measures from structural MRI instead of voxel-based features.

5.1. Comparisons with other studies

Several studies have investigated neuroimaging techniques for the early detection of AD, with a focus on conversion in MCI subjects and separating AD patients from HCs. In one study [65], researchers proposed a multitask learning method based on multimodality data such as MRI, fluorodeoxyglucose positron emission tomography (FDG-PET), and CSF. They reported an accuracy of 93.30% and 73.9% on AD/HC and pMCI/sMCI samples, respectively. One study [73] used cortical thickness data with the manifold harmonic transform method using MRI data. They achieved an SEN of 82% and an SPE of 93% on AD/HC, and an SEN of 63% and an SPE of 76% on sMCI and pMCI samples, respectively. Another study [74] presented a morphological factor method with an ACC

Table 6
Classification of Alzheimer's disease and healthy control subjects.

Author	Data	AD/HC	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Zhang et al., 2012 [65]	MRI	45/50	10 Fold	84.80	–	–	–
Zhang et al., 2012 [65]	MRI + CSF + PET	45/50	10 Fold	93.30	–	–	–
Cho et al., 2012 [73]	MRI	128/160		–	82	93	–
Liu et al., 2014 [67]	MRI+FDG-PET	51/52	10 Fold	94.37	94.71	94.04	97.24
Westman et al., 2012 [8]	MRI	96/111	10 Fold	87	83.3	90.1	93.00
Westman et al., 2012 [8]	MRI + CSF	96/111	10 Fold	91.8	88.5	94.6	95.80
Hu et al., 2016 [63]	MRI	228/188	Leave-one-out	84.13	82.45	85.63	90.00
Proposed method	**MRI**	**92/94**	**10 Fold**	**93.01**	**89.13**	**96.80**	**93.51**

Table 7
Classification of pMCI and sMCI subjects.

Author	Data	pMCI/ sMCI	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC (%)	Conversion time (months)
Zhang et al., 2012 [65]	MRI	43/48	10 Fold	62.00	56.60	60.20	–	0–24
Zhang et al., 2012 [65]	MRI + CSF + PET	43/48	10 Fold	73.90	68.80	73.60	–	0–24
Cho et al., 2012 [73]	MRI	72/131		–	63	76	–	0–18
Aksu et al., 2011 [74]	MRI	20/29		72.3				0–36
Young et al., 2013 [68]	MRI	47/96		61.5	–	–	–	0–36
Young et al., 2013 [68]	MRI + PET + CSF + APOE	47/96		74.1	–	–	–	0–36
Liu et al.,2014 [67]	MRI+FDG-PET	43/56	10 Fold	67.83	64.88	70.00	69.57	0–18
Westman et al., 2012 [8]	MRI	81/81	10 Fold	65.40	65.40	65.40	73.10	0–36
Westman et al., 2012 [8]	MRI + CSF	81/81	10 Fold	68.5	74.1	63	76.00	0–36
Cheng et al., 2015 [46]	MRI	43/56	10 Fold	73.4	74.3	72.1	76.40	0–24
Cheng et al., 2015 [46]	MRI+CSF+PET	43/56	10 Fold	79.4	84.5	72.7	84.80	**0–24**
Moradi et al., 2015 [10]	MRI	164/100	10 Fold	74.74	88.85	51.59	76.61	**0–36**
Moradi et al., 2015 [10]	MRI+ age and cognitive measures	164/100	10 Fold	81.72	86.65	73.64	90.20	**0–36**
Davatzikos et al., 2010 [66]	MRI+CSF	69/170	5 Fold	61.70	84.20	51.20	67.70	**0–12**
Hu et al., 2016 [63]	MRI	71/62	Leave-one-out	76.69	71.83	82.26	79.00	**0–36**
Proposed method	**MRI**	71/65	10 Fold	75.00	76.92	73.23	75.08	**0–36**

of 72.3% from MRI data on sMCI and pMCI subjects. In that study [68], the authors investigated a Gaussian process approach using several combined multimodality data sources (i.e., MRI, PET, CSF, and APOE genotypes). They reported a balanced ACC of 74.1% on sMCI/pMCI groups. Another study [67] used multimodality data (MRI and PET) with a multitask feature-selection method with an accuracy of 94.37% and 67.83% on AD/HC and pMCI/sMCI samples, respectively. In one study [66], the researchers combined statistical analysis and pattern classification methods by using multimodality data (MRI and CSF) with an ACC of 61.7% on pMCI/sMCI samples. Another study [8] investigated a multivariate data analysis method using multimodality data (i.e., MRI and CSF). The researchers reported an ACC of 68.5% on pMCI/sMCI data and an ACC of 91.8% on AD and HC data. In their study [46], the authors introduced domain transfer learning using multimodality data (i.e., MRI, CSF and PET), with an ACC of 79.4% on pMCI/sMCI samples. In another study [10], the authors employed VBM analysis of GM as a feature, combining age and cognitive measures. They reported an ACC of 82% on pMCI/sMCI samples. In another study [63], the authors presented a multiscale feature extraction from baseline MRI data. They reported an ACC of 84.13% in the classification of AD/HC and 76.69% for prediction of pMCI/sMCI. Tables 6, 7 show the detailed parameters of classification performance using various methods for the AD/HC and pMCI/sMCI classifications, which were the main tasks in the current study. As can be seen in Tables 6, 7, the performance of the proposed system was highly competitive for the performance terms when compared to the other systems reported in the literature for AD classification and the prediction of MCI-to-AD conversion.

6. Conclusion

In this paper, we proposed a novel and automatic feature-selection method, based on feature ranking and a GA, to select the optimal features with maximum discriminative and minimum numbers of selected features for MCI conversion prediction and AD classification. In the proposed method, we used the Fisher criterion as part of the objective function in the GA in order to evaluate the feature subsets. The proposed feature-selection method not only selects the top discriminative features but also minimizes the dimensionality of the input vectors to low-dimensional space. The experimental results demonstrate that a combination of feature ranking and GA is a reliable technique for MCI conversion prediction and early detection of AD, especially with regard to high-dimensional data pattern recognition. In addition, the experimental results show that the GM atrophy pattern in AD subjects and HCs, as shown in VBM analysis, can be useful for extracting features from sMCI and pMCI samples. The performance of the proposed CAD system was measured on 458 subjects from the ADNI dataset using 10-fold cross validation. The experimental results show that the performance of the proposed approach can compete strongly with state-of-the-art techniques using MRI data, as reported in the literature.

Acknowledgment

This work was partly carried out under the Brain Mapping by Integrated Neuro technologies for Disease Studies (Brain/MINDS) project (grant number16dm0207017h0003), funded by the Japan Agency for Medical Research and Development (AMED).

References

[1] Alzheimer's Association | Alzheimer's Disease and Dementia, 2015. ⟨http://www.alz.org/⟩ (Accessed 5 April 2015).
[2] A. Association, Alzheimer' s Association Report 2015 Alzheimer' s disease facts and figures, Alzheimer'S. Dement. 11 (2015) 332–384. http://dx.doi.org/10.1016/j.jalz.2015.02.003.
[3] E.E. Tripoliti, D.I. Fotiadis, M. Argyropoulou, G. Manis, A six stage approach for

the diagnosis of the Alzheimer's disease based on fMRI data, J. Biomed. Inf. 43 (2010) 307–320. http://dx.doi.org/10.1016/j.jbi.2009.10.004.
[4] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A.D.N. Initiative, et al., 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.
[5] 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.
[6] I. Beheshti, H. Demirel, Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease, Comput. Biol. Med. 64 (2015) 208–216. http://dx.doi.org/10.1016/j.compbiomed.2015.07.006.
[7] C. Aguilar, E. Westman, J.-S.S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, et al., Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment (A. Simmons, L.O. Wahlund)Psychiatry Res. Neuroimaging 212 (2013) 89–98. http://dx.doi.org/10.1016/j.pscychresns.2012.11.005.
[8] 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.
[9] M. Li, Y. Qin, F. Gao, W. Zhu, X. He, Discriminative analysis of multivariate features from structural MRI and diffusion tensor images, Magn. Reson. Imaging 32 (2014) 1043–1051. http://dx.doi.org/10.1016/j.mri.2014.05.008.
[10] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A.D.N. Initiative, et al., Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398–412. http://dx.doi.org/10.1016/j.neuroimage.2014.10.002.
[11] E.E. Bron, M. Smits, W.M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J.M. Papma, R.M.E. Steketee, C. Méndez Orellana, R. Meijboom, M. Pinto, J.R. Meireles, C. Garrett, A.J. Bastos-Leite, A. Abdulkadir, O. Ronneberger, N. Amoroso, R. Bellotti, D. Cárdenas-Peña, A.M. Álvarez-Meza, C.V. Dolph, K.M. Iftekharuddin, S.F. Eskildsen, P. Coupé, V.S. Fonov, K. Franke, C. Gaser, C. Ledig, R. Guerrero, T. Tong, K.R. Gray, E. Moradi, J. Tohka, A. Routier, S. Durrleman, A. Sarica, G. Di Fatta, F. Sensi, A. Chincarini, G.M. Smith, Z.V. Stoyanov, L. Sørensen, M. Nielsen, S. Tangaro, P. Inglese, C. Wachinger, M. Reuter, J.C. van Swieten, W.J. Niessen, S. Klein, Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge, Neuroimage 111 (2015) 562–579. http://dx.doi.org/10.1016/j.neuroimage.2015.01.048.
[12] A.H. Andersen, W.S. Rayens, Y. Liu, C.D. Smith, Partial least squares for discrimination in fMRI data, Magn. Reson. Imaging 30 (2012) 446–452. http://dx.doi.org/10.1016/j.mri.2011.11.001.
[13] Y. Fan, S.M. Resnick, X. Wu, C. Davatzikos, Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study, Neuroimage 41 (2008) 277–285.
[14] E. Dinesh, M.S. Kumar, M. Vigneshwar, T. Mohanraj, Instinctive classification of Alzheimer's disease using FMRI, pet and SPECT images, in: Intell. Syst. Control (ISCO), 2013 7th International Conference, pp. 405–409. http://dx.doi.org/10.1109/ISCO.2013.6481189, 2013.
[15] L. Mesrob, DTI and structural MRI classification in Alzheimer's Disease, Adv. Mol. Imaging 2 (2012) 12–20. http://dx.doi.org/10.4236/ami.2012.22003.
[16] M. Graña, M. Termenon, a. Savio, a. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, a. Besga, Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation, Neurosci. Lett. 502 (2011) 225–229. http://dx.doi.org/10.1016/j.neulet.2011.07.049.
[17] 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.compbiomed.2013.07.004.
[18] H. Hanyu, T. Sato, K. Hirao, H. Kanetaka, T. Iwamoto, K. Koizumi, The progression of cognitive deterioration and regional cerebral blood flow patterns in Alzheimer's disease: a longitudinal SPECT study, J. Neurol. Sci. 290 (2010) 96–101. http://dx.doi.org/10.1016/j.jns.2009.10.022.
[19] K.R. Gray, R. Wolz, R. a. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, A.D.N. Initiative, et al., Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease, Neuroimage 60 (2012) 221–229. http://dx.doi.org/10.1016/j.neuroimage.2011.12.071.
[20] Y.J. Chen, G. Deutsch, R. Satya, H.-G.G. Liu, J.M. Mountz, A semi-quantitative method for correlating brain disease groups with normal controls using SPECT: alzheimer's disease versus vascular dementia, Comput. Med. Imaging Graph 37 (2013) 40–47. http://dx.doi.org/10.1016/j.compmedimag.2012.11.001.
[21] J.M.M. Górriz, F. Segovia, J. Ramírez, A. Lassl, D. Salas-Gonzalez, GMM based SPECT image classification for the diagnosis of Alzheimer's disease, Appl. Soft Comput. 11 (2011) 2313–2325. http://dx.doi.org/10.1016/j.asoc.2010.08.012.
[22] C.H.C. Huang, B.Y. Bin Yan, H.J.H. Jiang, D.W.D. Wang, Combining Voxel-based Morphometry with Artifical Neural Network Theory in the Application Research of Diagnosing Alzheimer's Disease, in: Proceedings of the 2008 International Conference on Biomed. Eng. Informatics, 1 pp. 250–254. doi:10.1109/BMEI.2008.245, 2008.
[23] 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: Proceedings of the 2008 Fourth International Conference on Nat. Comput., 2 pp. 197–201. doi:10.1109/ICNC.2008.804, 2008.
[24] A. Savio, M.T. García-Sebastián, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga, A. López, de Munain, J. Villanúa, 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.compbiomed.2011.05.010.
[25] A. Chincarini, P. Bosco, P. Calvini, G. Gemme, M. Esposito, C. Olivieri, L. Rei, S. Squarcia, G. Rodriguez, R. Bellotti, P. Cerello, I. De Mitri, A. Retico, F. Nobili, et al., Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease, Neuroimage 58 (2011) 469–480. http://dx.doi.org/10.1016/j.neuroimage.2011.05.083.
[26] E. Westman, L. Cavallin, J.-S.S. Muehlboeck, Y. Zhang, P. Mecocci, B. Vellas, M. Tsolaki, I. Kłoszewska, H. Soininen, C. Spenger, S. Lovestone, A. Simmons, L.O. Wahlund, et al., Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease, PLoS One 6 (2011) e22506. http://dx.doi.org/10.1371/journal.pone.0022506.
[27] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, G. Catheline, Classification of Alzheimer's disease subjects from MRI using hippocampal visual features, Multimed. Tools Appl 74 (2014) 1249–1266. http://dx.doi.org/10.1007/s11042-014-2123-y.
[28] S. Li, F. Shi, F. Pu, X. Li, T. Jiang, S. Xie, Y. Wang, Hippocampal shape analysis of Alzheimer disease based on machine learning methods, Am. J. Neuroradiol. 28 (2007) 1339–1345. http://dx.doi.org/10.3174/ajnr.A0620.
[29] P. Coupé, S.F. Eskildsen, J.V. Manjón, V.S. Fonov, D.L. Collins, A. disease Neuroimaging Initiative, et al., Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease, Neuroimage 59 (2012) 3736–3747. http://dx.doi.org/10.1016/j.neuroimage.2011.10.080.
[30] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S.S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero, et al., Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging (F. Eustache, O. Colliot), Neuroimage 47 (2009) 1476–1486. http://dx.doi.org/10.1016/j.neuroimage.2009.05.036.
[31] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, A.D.N.A.D.N. Initiative, et al., Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI, Hippocampus 19 (2009) 579. http://dx.doi.org/10.1002/hipo.20626.
[32] 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.
[33] 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.
[34] 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.
[35] S. Klöppel, C.M. Stonnington, C. Chu, B. Draganski, R.I. Scahill, J.D. Rohrer, N.C. Fox, C.R. Jack, J. Ashburner, R.S.J. Frackowiak, Automatic classification of MR scans in Alzheimer's disease, Brain 131 (2008) 681–689. http://dx.doi.org/10.1093/brain/awm319.
[36] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. Van Der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl, A. Dale, Reliability in multi-site structural MRI studies : effects of gradient non-linearity correction on phantom and human data, Neuroimage 30 (2006) 436–443. http://dx.doi.org/10.1016/j.neuroimage.2005.09.046.
[37] J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, Med. Imaging IEEE Trans. 17 (1998) 87–97.
[38] I. Beheshti, H. Demirel, Feature-ranking-based Alzheimer's disease classification from structural MRI, Magn. Reson. Imaging 34 (2016) 252–263. http://dx.doi.org/10.1016/j.mri.2015.11.009.
[39] H. Matsuda, S. Mizumura, K. Nemoto, F. Yamashita, E. Imabayashi, N. Sato, T. Asada, 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.
[40] Y. Hirata, H. Matsuda, K. Nemoto, T. Ohnishi, K. Hirao, F. Yamashita, T. Asada, S. Iwabuchi, H. Samejima, 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.
[41] X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, L. Yao, K. Chen, Voxel-based assessment of gray and white matter volumes in Alzheimer's disease, Neurosci. Lett. 468 (2010) 146–150. http://dx.doi.org/10.1016/j.neulet.2009.10.086.
[42] 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.
[43] G.B. Karas, E.J. Burton, S. a.R.B. Rombouts, R. a. Van Schijndel, J.T. O'Brien, P. Scheltens, I.G. McKeith, D. Williams, C. Ballard, F. Barkhof, 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.
[44] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (2007) 95–113. http://dx.doi.org/10.1016/j.neuroimage.2007.07.007.
[45] A. Klein, J. Andersson, B. a. Ardekani, J. Ashburner, B. Avants, M.C. Chiang,

G.E. Christensen, D.L. Collins, J. Gee, P. Hellier, J.H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R.P. Woods, J.J. Mann, R.V. Parsey, 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.

[46] B. Cheng, M. Liu, D. Zhang, B.C. Munsell, Domain Transfer Learning for MCI Conversion Prediction, IEEE Trans. Biomed. Eng. 62 (2015) 1805–1817.

[47] M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río, 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.

[48] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-gonzalez, M.M. López, F. Segovia, R. Chaves, , F-FDG PET imaging analysis for computer aided Alzheimer' s diagnosis 181 (2011) 903–916. http://dx.doi.org/10.1016/j.ins.2010.10.027.

[49] N. Zhou, L. Wang, A modified *t*-test feature selection method and its application on the HapMap genotype dData, Genom. Proteom. Bioinformatics 5 (2007) 242–249. http://dx.doi.org/10.1016/S1672-0229(08)60011-X.

[50] J. Pohjalainen, O. Räsänen, S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Comput. Speech Lang. 29 (2015) 145–171. http://dx.doi.org/10.1016/j.csl.2013.11.004.

[51] W. Duch, T. Wieczorek, J. Biesiada, M. Blachnik, Comparison of feature ranking methods based on information entropy, in: IEEE International Conference on Neural Networks - Conference Proceedings 2 1415–1419. doi:10.1109/IJCNN.2004.1380157, 2004.

[52] W. Yan, Fusion in multi-criterion feature ranking, in: Proceedings of Inf. Fusion, 2007 10th International Conference 1–6. doi:10.1109/ICIF.2007.4408064, 2007.

[53] X. Geng, T. Liu, T. Qin, H. Li, Feature selection for ranking, Sigir (2007) 407–414.

[54] R. Ruiz, J. Riquelme, J. Aguilar-Ruiz, Fast feature ranking algorithm, Knowl.-Based Intell. (2003) 325–331 ⟨http://link.springer.com/chapter/10.1007/978-3-540-45224-9_46⟩.

[55] Y.-W. Chang, C.-J. Lin, Feature ranking using linear svm, in: JMLRWorkshop Confrence Proceedings, 3, pp. 53–64, 2008.

[56] I. Slavkov, B. Zenko, S. Dzeroski, Evaluation method for feature rankings and their aggregations for biomarker discovery, J. Mach. Learn. Res. - Proc. Track. 8 (2010) 122–135 ⟨http://dblp.uni-trier.de/db/journals/jmlr/jmlrp8.html#SlavkovZD10⟩.

[57] M. Liu, D. Zhang, D. Shen, Ensemble sparse classification of Alzheimer's disease, Neuroimage 60 (2012) 1106–1116. http://dx.doi.org/10.1016/j.neuroimage.2012.01.055.

[58] R. Chaves, J. Ramírez, J.M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SVM-based computer-aided diagnosis of the Alzheimer's disease using *t*-test NMSE feature selection with feature correlation weighting, Neurosci. Lett. 461 (2009) 293–297. http://dx.doi.org/10.1016/j.neulet.2009.06.052.

[59] D. Wang, H. Zhang, R. Liu, W. Lv, D. Wang, T-Test feature selection approach based on term frequency for text categorization, Pattern Recognit. Lett. 45 (2014) 1–10. http://dx.doi.org/10.1016/j.patrec.2014.02.013.

[60] 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.

[62] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory, Clin. Neurophysiol. 126 (2015) 2132–2141. http://dx.doi.org/10.1016/j.clinph.2015.02.060.

[63] K. Hu, Y. Wang, K. Chen, L. Hou, X. Zhang, Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis, Neurocomputing 175 (2016) 132–145. http://dx.doi.org/10.1016/j.neucom.2015.10.043.

[64] G. Prasad, S.H. Joshi, T.M. Nir, A.W. Toga, P.M. Thompson, Brain connectivity and novel network measures for Alzheimer's disease classification, Neurobiol. Aging 36 (2015) S121–S131. http://dx.doi.org/10.1016/j.neurobiolaging.2014.04.037.

[65] D. Zhang, D. Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, Neuroimage 59 (2012) 895–907. http://dx.doi.org/10.1016/j.neuroimage.2011.09.069.

[66] C. Davatzikos, P. Bhatt, L.M. Shaw, K.N. Batmanghelich, J.Q. Trojanowski, Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification, Neurobiol. Aging 32 (2011). http://dx.doi.org/10.1016/j.neurobiolaging.2010.05.023.

[67] F. Liu, C. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer â€™ s Disease and mild cognitive impairment identification, Neuroimage 84 (2014) 466–475. http://dx.doi.org/10.1016/j.neuroimage.2013.09.015.

[68] J. Young, M. Modat, M.J. Cardoso, A. Mendelson, D. Cash, S. Ourselin, N. Initiative, Accurate multimodal probabilistic prediction of conversion to Alzheimer â€™ s disease in patients with mild cognitive impairment, NeuroImage Clin. 2 (2013) 735–745. http://dx.doi.org/10.1016/j.nicl.2013.05.004.

[69] C. Davatzikos, F. Xu, Y. An, Y. Fan, S.M. Resnick, Longitudinal progression of Alzheimers-like patterns of atrophy in normal older adults: the SPARE-AD index, Brain 132 (2009) 2026–2035. http://dx.doi.org/10.1093/brain/awp091.

[70] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G.B. Frisoni, D.L. Collins, MRI-based automated computer classification of probable AD versus normal controls, IEEE Trans. Med. Imaging 27 (2008) 509–520. http://dx.doi.org/10.1109/TMI.2007.908685.

[71] Y. Fan, D. Shen, C. Davatzikos, Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM, Med. Image Comput. Comput. Assist. Interv. 8 (2005) 1–8. http://dx.doi.org/10.1007/11566465_1.

[72] Y. Fan, N. Batmanghelich, C.M. Clark, C. Davatzikos, Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, Neuroimage 39 (2008) 1731–1743. http://dx.doi.org/10.1016/j.neuroimage.2007.10.031.

[73] Y. Cho, J. Seong, Y. Jeong, S. Yong, Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data, Neuroimage 59 (2012) 2217–2230. http://dx.doi.org/10.1016/j.neuroimage.2011.09.085.

[74] Y. Aksu, D.J. Miller, G. Kesidis, D.C. Bigler, Q.X. Yang, An MRI-derived definition of MCI-to-AD conversion for long-term, automatic prognosis of MCI patients, PLoS One 6 (2011). http://dx.doi.org/10.1371/journal.pone.0025074.