COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 137 (2016) 177–193

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error

Iman Beheshti [a,*], *Hasan Demirel* [b], *Farnaz Farokhian* [c], *Chunlan Yang* [c], *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 Lab, Department of Electrical & Electronic Engineering, Eastern Mediterranean University, Famagusta, Mersin 10, Turkey*
[c] *College of Life Science and Bioengineering, Beijing University of Technology, Beijing, 100022, China*

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form
2 September 2016
Accepted 22 September 2016

Keywords:
Alzheimer's disease
Structural MRI
Feature extraction
Feature ranking
Computer-aided diagnosis
Classification error

ABSTRACT

Background and objective: This paper presents an automatic computer-aided diagnosis (CAD) system based on feature ranking for detection of Alzheimer's disease (AD) using structural magnetic resonance imaging (sMRI) data.
Methods: The proposed CAD system is composed of four systematic stages. First, global and local differences in the gray matter (GM) of AD patients compared to the GM of healthy controls (HCs) are analyzed using a voxel-based morphometry technique. The aim is to identify significant local differences in the volume of GM as volumes of interests (VOIs). Second, the voxel intensity values of the VOIs are extracted as raw features. Third, the raw features are ranked using a seven-feature ranking method, namely, statistical dependency (SD), mutual information (MI), information gain (IG), Pearson's correlation coefficient (PCC), t-test score (TS), Fisher's criterion (FC), and the Gini index (GI). The features with higher scores are more discriminative. To determine the number of top features, the estimated classification error based on training set made up of the AD and HC groups is calculated, with the vector size that minimized this error selected as the top discriminative feature. Fourth, the classification is performed using a support vector machine (SVM). In addition, a data fusion approach among feature ranking methods is introduced to improve the classification performance.
Results: The proposed method is evaluated using a data-set from ADNI (130 AD and 130 HC) with 10-fold cross-validation. The classification accuracy of the proposed automatic system for the diagnosis of AD is up to 92.48% using the sMRI data.
Conclusions: An automatic CAD system for the classification of AD based on feature-ranking method and classification errors is proposed. In this regard, seven-feature ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI) are evaluated. The optimal size of top discriminative features is determined by the classification error estimation in the training phase. The experimental results indicate that the performance of the proposed system is comparative to that of state-of-the-art classification models.

* *Corresponding author.* Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo, 187-8551, Japan.
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.cmpb.2016.09.019
0169-2607/

1. Introduction

Alzheimer's disease (AD), a progressive irreversible neurodegenerative disorder, occurs most frequently in older adults and gradually destroys regions of the brain that are responsible for memory, thinking, learning, and behavior [1]. It is estimated that 5.3 million Americans of all ages suffer from AD in 2015 [2]. Among the top 10 causes of death among Americans, AD is the only disease that cannot be cured, prevented, or slowed [2]. Although there is no cure for AD, early detection may shed light on AD mechanisms and improve the responses of AD patients to drug therapy and their quality of life. In recent years, the analysis of neuroimaging data, such as structural magnetic resonance imaging (sMRI) [3–12], functional MRI [13–15], and diffusion tensor imaging [16–18], in addition to positron emission tomography (PET) and single photon emission computed tomography (SPECT) [19–24], has attracted much interest, with recent improvements in accurate detection of AD. In this paper, we focus only on the use of sMRI data in the classification of AD. Recently, sMRI brain data have been widely used to design computer-aided diagnosis (CAD) systems for the classification of AD [4,9,25,26], because of the noninvasiveness, excellent spatial resolution, and good tissue contrast of sMRI, in addition to the absence of radioactive pharmaceutical injection, as occurs with PET and SPECT [19–22]. Many researchers studied advanced pattern analysis and classification approaches for extracting complex spatial patterns of brain structure [14,27–30]. This paper describes the application of an automatic CAD system, which uses statistical feature-ranking methods as part of a novel feature-selection process, followed by estimation of the classification error in AD and healthy control (HC) groups to determine the optimum number of highest-ranking features to be selected. In the training set, resubstitution and cross-validation error estimators were used as classification errors to measure the quality of a classifier. We used these classification error metrics as stopping criteria among the ranked features to estimate the optimal number of features with the most discriminative information in the classification process. We evaluated seven feature-ranking methods, namely, statistical dependency (SD), mutual information (MI), information gain(IG), Pearson's correlation coefficient (PCC), the *t*-test score (TS), Fisher's criterion (FC), and the Gini index (GI) in the proposed CAD system. In the proposed approach, high-dimensional feature space was reduced into lower dimensional space by employing the minimized classification error as the dimensionality selection criterion in an iterative process of incrementing the number of ranked features. The proposed feature-selection method was applied to gray matter (GM) atrophy clusters of voxels, which corresponded to the volume of interests (VOIs) of the sMRI data obtained through the voxel-based morphometry (VBM) analysis during preprocessing. VBM is an advanced method used to assess the whole-brain structure using voxel-by-voxel comparisons [8,31–36]. It is one of the best methods for feature extraction from sMRI in AD [9]. In the proposed system, we used only sMRI data. The proposed CAD system was applied in four stages in a systematic manner. In the first stage, the VBM technique was employed, in addition to diffeomorphic anatomical registration using the exponentiated Lie algebra (DARTEL) [33]. This approach was used to analyze group-wise comparisons between cross-sectional structural MRI scans to detect the MRI voxels that were best discriminated between the AD group versus HCs [8,31–33]. Based on the VBM and DARTEL approach on a global brain scale, and regional structural GM alterations, regions with significant atrophy of GM were investigated and specified in the patients who suffer from AD. In the second stage, specified VOIs were used as 3D masks for extracting voxel intensity values from GM atrophy regions to generate raw feature vectors. These feature vectors were subjected to further data-selection processes before they were used by the classifier. In the third stage, the extracted features were ranked based on the statistical scores (i.e., SD, MI, IG, PCC, TS, FC, and GI) of the AD and HC groups in the training set. The ranking scores can be considered an indicator of the level of separation/discrimination between the AD and HC groups in the training set. Feature ranking has been used successfully in a number of pattern-recognition studies [37–42]. In addition, an automatic approach based on classification error estimation was used to determine the number of top features using the AD and HC groups in the training set. This approach adaptively determines the optimum number of top features and identifies a discriminative subset of high-performance features based on the training data in each fold instead of using a fixed number of features. In the fourth stage, the performance of the proposed feature-selection technique was evaluated using a support vector machine (SVM) classifier. In this work, the SVM classifier with a linear kernel was trained to discriminate between the classes. In addition, instead of using a single feature ranking method, the results of multiple individual feature ranking methods were combined through the proposed data fusion technique for improved classification performance.

In summary, the aim of this study was to design an automatic CAD system based on statistical feature ranking and classification errors as part of a novel feature-selection method. The proposed system utilizes feature ranking based on statistical scores, followed by the determination of resubstitution and cross-validation error estimators to identify the number of ranked features that minimizes the error in the training set. This process helps to identify a selected discriminative subset of high-performance features into a lower-dimensional feature vector space representing sMRI images. In addition, a data fusion technique was proposed to improve the AD classification performance among different feature ranking methods. The performance of the proposed system was assessed using a data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) containing 260 subjects (130 AD patients and 130 HCs) using 10-fold cross-validation. The experimental results showed that the accuracy (ACC) (92.48%), sensitivity (SEN) (91.07%), specificity (SPE) (93.89%), and area under the curve (AUC) (0.963) of the proposed system were well comparatively to results obtained with state-of-the-art techniques in terms of AD classification.

The rest of the paper is organized as follows: Section 2 details the statistical data in the study. Section 3 describes the proposed methodology to design an automatic CAD system based on feature ranking and classification error. Section 4 presents the experimental results, discussion, and analysis of the proposed system. Finally, Section 5 presents the conclusions.

2. Materials

2.1. MRI acquisition

The MR images and data used in this study were obtained from the ADNI database.[1] All the participants initially underwent a number of neuropsychological examinations, resulting in several clinical characteristic indicators, including the Mini Mental State Examination (MMSE) score and Clinical Dementia Ratio (CDR) score. The MRI scans were acquired using 3 Tesla, T1-weighted by Siemens scanner with Acquisition Plane = SAGITTAL, Acquisition Type = 3D, Coil = PA, Flip Angle = 9.0 degree, Matrix X/Y/Z = 240.0 pixels /256 pixels /176 pixels, Mfg Model = Skyra, Pixel Spacing X/Y = 1.0 mm/1.0 mm, Pulse Sequence = GR/IR, Slice Thickness = 1.2 mm, TE = 2.98 ms, TI = 900 ms and TR = 2300 ms. Additional MR image corrections were also applied such as *GradWarp* for correction of geometric distortion caused by gradient non-linearity [43] and *N3* bias field correction for reducing intensity inhomogeneity due to non-uniformities in the radio frequency receiver coils [44] .

2.2. Subjects

The HC group contained 130 participants, with ages ranging from 56 to 88 years (mean 74.49 ± 6.13 years), MMSE scores ranging from 27 to 30 (mean of 29.26 ± 0.80), and a CDR score of zero. The AD group contained 130 patients, and their ages ranged from 57 to 91 years (mean of 75.88 ± 7.54 years). Their MMSE and CDR scores ranged from 10 to 28 (mean of 22.33 ± 3.27), and 0.5 to 2 (mean of 0.80 ± 0.37).

3. Proposed CAD classification system

In this section, an automatic CAD system, which is based on feature ranking, followed by optimal selection of a number of top features using a classification error for high-performance AD classification, is introduced. An outline of the proposed ranking-based CAD system is illustrated in Fig. 1. First, the VBM and DARTEL approach were employed to preprocess 3D T1-weighted MRI data. Second, voxel-based feature extraction was performed. Third, the extracted features were ranked based on the score values of the respective features in the training set. The optimal number of top ranked features was automatically obtained by minimizing the classification error among the possible number of features. These approaches resulted in high-dimensional sMRI data of VOI in a low-dimensional space with a discriminative subset of high-performance features based on the training data in each fold. Fourth, to evaluate the performance of the proposed feature-selection method, a linear SVM classifier was employed. In addition, a data fusion technique among different feature ranking methods was engaged to improve the classification performance.

3.1. MRI data preprocessing

The 3D T1-weighted brain images were pre-processed using the SPM8 package[2] and VBM8 toolbox.[3] Recently, several studies have been used VBM method for detection atrophic changes in AD [8,9,33,45–47]. In this study, DARTEL approach was employed with VBM to increase enhancement of inter-subject registration provide precise, accurate localization of structural damage of the MRI images. DARTEL template is generated from 550 healthy control participants (defined by default setting of VBM8 toolbox) [48]. In the VBM8 toolbox, all the sMRI data were bias-corrected and segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) components. The normalized segmented images were modulated by applying a nonlinear deformation. This allows the comparison of absolute amounts of tissue corrected for individual differences in brain size [48]. The deformation is applied to segmented images to create an image which is in voxel-for-voxel registration with the template [49]. In the current study, we used only GM component. Finally, the all GM components were spatially smoothed with an 8 mm full-width-half-maximum Gaussian smoothing kernel. After spatial preprocessing, the smoothed, modulated, DARTEL-warped and normalized GM datasets were subjected to the general linear model to detect gray matter volume changes using voxel-wise two sample t-test in SPM8. Age was engaged into the matrix design as a nuisance variable. The whole brain analysis was implemented using a p-value of <0.01 with correction for family-wise error (FWE). The extent threshold was adjusted at 1400 voxels for two-sample comparisons. Regional changes in GM volumes were detected by a voxel-based analysis of the entire brain. The overall procedure is explained in more detail in Ref. [25].

3.2. Feature extraction

The brain regions containing significantly decreased GM volumes were obtained using the VBM plus DARTEL analysis in the AD patients relative to the HCs. Based on the VBM and DARTEL results, a 3D mask generated from atrophy regions was modeled to identify VOIs for further processing. This mask was applied to the GM density volumes resulting from the VBM and DARTEL analyses to extract voxels from atrophy regions as raw feature vectors. The subjects were randomly divided into 10 folds, with the same number of AD and HC subjects in each fold. In each iteration, one of the folds was used for testing, and nine of the folds were used for training. A VBM analysis of each training data set was performed using a FWE corrected at $p < 0.01$ and extend threshold 1400 voxels to reveal regions of decreased GM volume in the patients through a 3D mask for the MRI samples in the respective training fold. In total, 10 different masks with different number of voxels (i.e., from 59,395 to 69,170 voxels) were defined. The respective 3D masks were used in the respective iteration to extract features from the training and testing data sets.

[1] www.loni.ucla.edu/ADNI.

[2] http://www.fil.ion.ucl.ac.uk/spm.

[3] http://dbm.neuro.uni-jena.de/vbm.

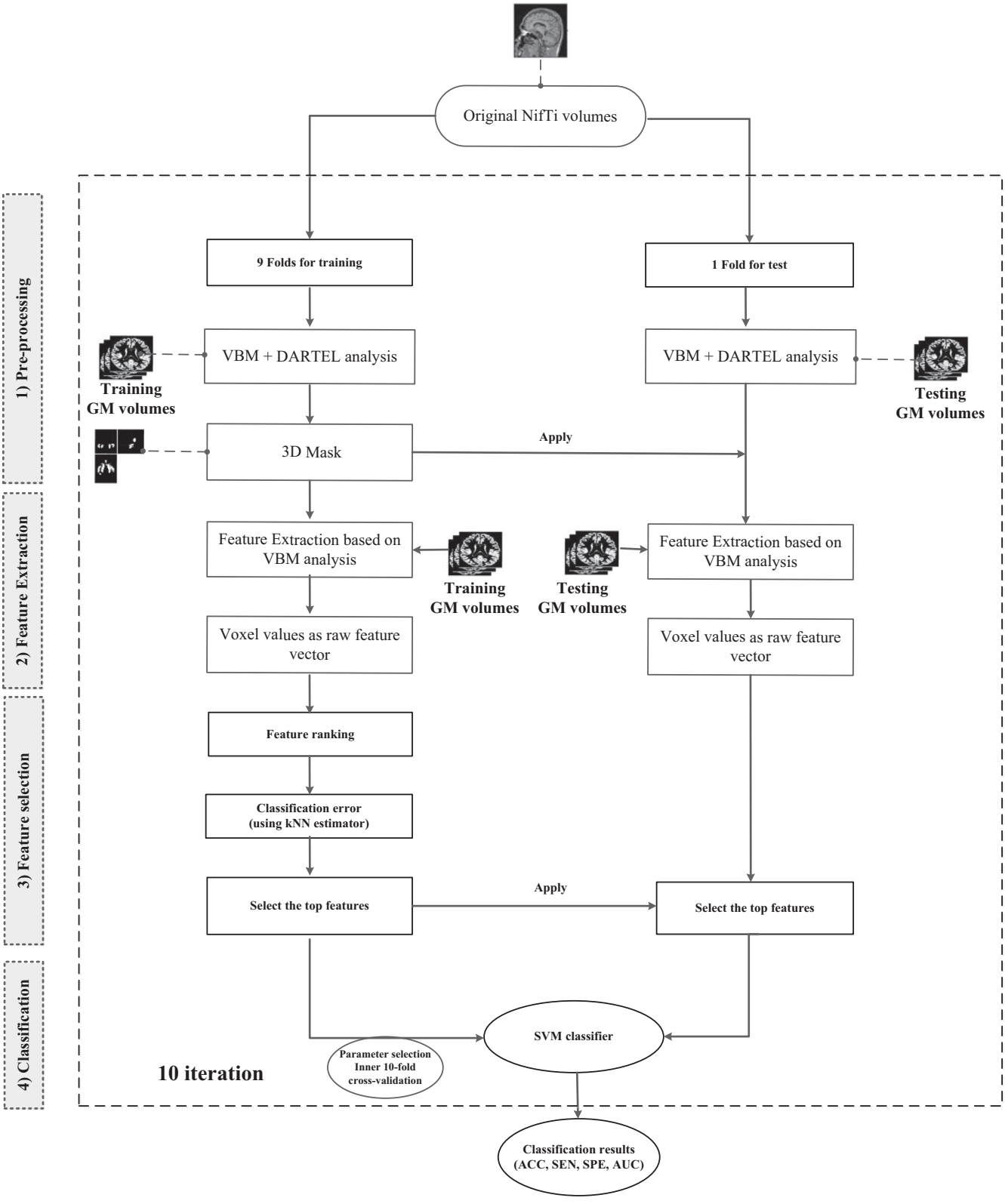

Fig. 1 – The pipeline of the proposed ranking-based CAD system for classifying AD.

3.3. *Proposed feature selection*

The dimensionality of raw feature spaces, which was very high, changed in line with the dimensionality of the 3D masks (i.e., from 59,395 to 69,170 voxels). It is expected that the feature vectors span a smaller region in the high-dimensional vector space. The aim of feature selection is to select the best features for improving the efficiency of learning, componential

cost and classification performance [50]. Feature selection using feature ranking is a reasonable approach to reduce the dimensionality and improve the performance, as the most discriminative subset of features is employed as the top feature representing the samples. Recently, several studies have used different feature ranking methods as part of feature selection in pattern recognition field [37,38,40–42,51,52].

3.3.1. *Feature ranking*

Feature ranking aids to achieve knowledge of data and identify relevant features and sort the features with respect to their relevance. On the other hand, feature ranking makes it easier to determine the relevance of features and class variables and to select the most informative/discriminative features, thereby improving the performance of classifier models and speeding up the learning process, especially when the dimensionality of a data setis very large [53]. Let $\Phi = [f^1, f^2, \ldots, f^M]$ be a features set containing M features, where the vector $f^j = (x_1^j, x_2^j, \ldots, x_N^j)^T$ is a vector of the values of a feature, f^j, N is the number of samples, and each value x_i^j of this vector shows a feature of that sample. A feature-ranking algorithm applied to data set Φ generates an ordered list of the features $\Psi = [f_*^1, f_*^2, \ldots, f_*^M]$. The superscript denotes the position in the ranked list of a feature, f_*, and the list is ordered by the reduction importance. Based on the feature ranking, we can select the top q-ranked features $[f_*^1, f_*^2, \ldots, f_*^q] q \le M$, where q can be determined by the user or adjusted experimentally [39]. In the present work, q was automatically estimated by minimizing the classification error of the training set in each fold.

In the present work, we used the following seven feature-ranking approaches. In each approach, the score of each feature was computed independently and sorted based on the respective score.

1. SD: SD measures the level of dependency between the values of a feature and the associated class labels. The SD between feature value X and class label C can be obtained as follows [51]:

$$SD = \sum_i \sum_j P(x_i, c_j) \frac{P(x_i, c_j)}{P(x_i)P(c_j)} \tag{1}$$

where $P(x_i, c_j)$ is the frequency count of data X with value x_i in the class c_j, $P(x_i)$ is the frequency count of data X with value x_i, and $P(c_j)$ is the frequency count of class C with value c_j. SD is nonnegative in the range of [0, 1], with SD = 0 indicating no correlation and SD = 1 denoting that C can be inferred once X is known. A larger SD means higher dependency between the feature value and class labels.

2. MI: MI measures the relevance of the feature value X and class label C by [51,54,55]:

$$MI = \sum_i \sum_j P(x_i, c_j) \log_2 \frac{P(x_i, c_j)}{P(x_i)P(c_j)} \tag{2}$$

MI is similar to SD. $P(x_i, c_j)$ is the frequency count of data X with value x_i in the class c_j, $P(x_i)$ is the frequency count of data X with value x_i, and $P(c_j)$ is the frequency count of class C with value c_j. MI is nonnegative in the range of [0, 1], with MI = 0 indicating no correlation, and MI = 1 meaning that C can be inferred once X is known.

3. IG: IG is a measure of the dependence between the features and class label. The IG of feature value X and class label C is calculated as follows [56]:

$$IG = H(X) - H(X|C) \tag{3}$$

Where $H(X)$ and $H(X|C)$ are the entropy of X and the entropy of X after observing C, respectively, as follows:

$$H(X) = -\sum_i P(x_i) \log_2 (P(x_i)) \tag{4}$$

$$H(X|C) = -\sum_j P(c_j) \sum_i P(x_i|c_j) \log_2 (P(x_i|c_j)) \tag{5}$$

The maximum value of IG is 1. Features with higher IG are more relevant.

4. PCC: PCC is a measure of the relevance between the features and class label. PCC of the feature value X and class label C is calculated as follows [54]:

$$PCC = \frac{\mathrm{cov}(X, C)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(C)}} \tag{6}$$

which in binary classification becomes:

$$PCC = \frac{\sum_{i=1}^{N} (x_i - \mu_x)(c_i - \mu_c)}{\sqrt{\sum_{i=1}^{N} (x_i - \mu_x)^2 (c_i - \mu_c)^2}} \tag{7}$$

where PCC is Pearson's correlation value, and μ_x and μ_c are the mean of all samples of X and C, respectively. PCC has a value in the range of [–1, 1]. PCC = 0 indicates independency of X and C, PCC = 1 denotes the highest positive correlation of them, and PCC = –1denotes the highest negative correlation. To select the top informative features, all the features were ranked according to their absolute PCC values.

5. TS: The TS measures the statistical significance of the value differences between the two classes. The t-test is performed by [57]:

$$TS = \frac{\mu_{c1} - \mu_{c2}}{\sqrt{\frac{\sigma_{c1}^2}{n_{c1}} + \frac{\sigma_{c2}^2}{n_{c2}}}} \tag{8}$$

Where TS is the t-test value and 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. To select the top informative features, all the features were ranked according to their absolute TS values.

6. FC: FC measures between-class and within-class scatter matrices between two classes, as shown below:

$$FC = \frac{w^T S_B w}{w^T S_W w} \tag{9}$$

where S_B and S_W represent the determinant of the between-class and within-class scatter matrices, respectively [58]. For two classes c_1 and c_2, the between-class scatter and within-class scatter matrixes are defined as follows:

$$S_B = (\mu_{c1} - \mu_{c2})(\mu_{c1} - \mu_{c2})^T \tag{10}$$

$$S_W = \sum_{x_i \in c_1} (x_i - \mu_{c1})(x_i - \mu_{c1})^T + \sum_{x_i \in c_2} (x_i - \mu_{c2})(x_i - \mu_{c2})^T \tag{11}$$

Where $w = S_W^{-1}(\mu_{c1} - \mu_{c2})$ and μ_{ci} are the mean of the data in each class.

To select the top informative features, all the features were ranked according to their FC values.

7. GI: The GI is a measure used to quantify the ability of a feature to distinguish between classes. The GI for a feature, f, is as follows [56]:

$$GI(f) = 1 - \sum_{i=1}^{C} [p(i|f]^2 \tag{12}$$

In the binary classification, the maximum value of the GI is 0.5, and features with a smaller GI are more relevant.

3.3.2. Classification error

Consider a labeled feature vector, $D = \{X, T\}$, where $X \in \Re^p$ (p is the dimension of the input vector) and C is the class label, which in binary classification with two classes $C \in \{-1, 1\}$. The pair $\{X, C\}$ has a joint probability distribution, F, which is unknown in practice. Let a classifier be trained with a set of n-independent observations, $S_n = \{(x_1, c_1), \ldots, (x_n, c_n)\}$, which are drawn from F. Let $\varphi : \{\Re^p \times \{-1, 1\}\}^n \times \Re^p \to \{-1, 1\}$ be a mapping input space to target as a classification rule, which maps S_n onto a classifier, $\varphi_n : \Re^p \to \{-1, 1\}$ [59]. The classification error e_n is the probability of an erroneous classification, which is calculated as follows [59,60]:

$$e_n = P(\varphi_n(X) \neq T | S_n) \tag{13}$$

In practice, the classification error is unknown, and the error must be estimated ($\hat{e}$). In the present work, two different classification error estimators were used: a resubstitution error ($\hat{e} = e_{resub}$) and across-validation error ($\hat{e} = e_{cross}$).

3.3.2.1. Resubstitution error. Consider a classifier, φ, which is trained with a set, $S_n = \{(x_1, c_1), \ldots, (x_n, c_n)\}$, where n is the number of samples. In the resubstitution error, e_{resub}, we design a classifier, S_n and test it on S_n to estimate the respective error, as follows:

$$e_{resub} = \frac{1}{n} \| (T_i - \varphi_n(x_i)) \|_0 \tag{14}$$

where $i = 1, \ldots, n$, and $\|v\|_0$ is the zero-norm counting the number of nonzero entries in v. The resubstitution estimator is nonrandomized, and it is very fast to compute in comparison to other error estimators, such as the cross-validation error estimator [61]. This estimator is always optimistically biased.

3.3.2.2. Cross-validation error estimator. The cross-validation error estimator is a randomized estimator obtained by randomly selecting K folds. In the K-fold cross-validation error estimator, the data are split into K folds at each step ($K = 10$): one fold is used as a test ($S_m = \{(x_1, c_1), \ldots, (x_m, c_m)\}$), and the remaining folds are used for training ($S_{n'} = \{(x_1, c_1), \ldots, (x_{n'}, c_{n'})\}$), where m and n' are the number of samples in the test and training sets, respectively. The above procedure is repeated K times by leaving a different fold as test data, which are used to compute (estimate) the classification error. In each iteration, the estimated respective error is calculated as follows:

$$e_K = \frac{1}{m} \| (T_i - \phi_{n'}(x_i)) \|_0 \tag{15}$$

where $i = 1, \ldots, m$. The total error was calculated using the average of the errors in each iteration.

$$e_{cross} = \frac{1}{K} \sum_{i=1}^{K} e_K \tag{16}$$

In this paper, we used the standard k-nearest-neighbors (k-NN) estimator to compute the classification error estimation. The k-NN estimator was chosen due to its lower computational cost relative to that of a state-of-the-art SVM estimator.

3.3.3. Optimal number of features based on the classification error

In addition to the feature-ranking algorithm based on the discriminative performance of the features, we propose to use an automatic approach based on classification error estimation to determine the number of top discriminative features and, hence, reduce the dimensionality of prospective feature vectors. Using this approach, it is simpler to automatically determine the q top discriminative features based on the ranked values in the training data in each fold instead of using a fixed q. Once the features were ranked, the number of top ranked features iteratively increased from 1 to $\Gamma (\Gamma \ll M)$ in the respective training error estimation. M is the number of features in the respective feature vectors in each fold, which had values from 59,395 to 69,170 voxels in our experiments. q is searched within the first Γ dimensions, where Γ is heuristically chosen to be 1500 to reduce the computational cost. Typical q values of between 10 and 1300 were observed in Γ. q is regarded as the optimal number of top ranked features that minimizes the classification error in the training set. The proposed algorithm to determine q, is given in the pseudo code shown in algorithm 1. The number of top features was iteratively incremented from 1 to Γ, using a training set of each fold to calculate the respective classification error estimation values by the k-NN estimator. Using a cross-validation process, the optimal numbers of top features, q, minimizing the classification error estimation in training phase was selected for use as the optimal dimension in the test and the training data in each

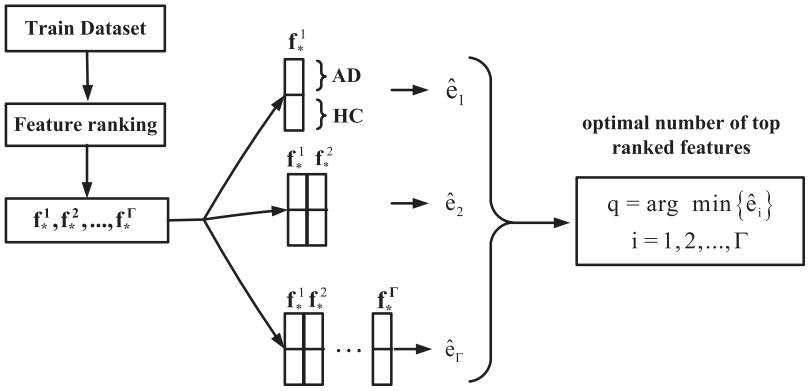

Fig. 2 – Detailed illustration of the proposed feature selection approach.

fold. Fig. 2 shows the details of the proposed feature selection procedure.

1: $V \leftarrow component\_set(Data_{Train}, Label_{Train})$
2: $Ranked\ features \leftarrow feature\_ranking(Data_{Train}, Label_{Train})$
2: $number\ of\ top\ features \leftarrow \emptyset,\ \Gamma = 1500$
3: **for** $n = 1$ *to* Γ **do**
4: $\hat{e}(n) \leftarrow Ranked\ features(1:n, Label_{Train})$
5: **end for**
6: $q \leftarrow \arg\min_{n \in 1,\ldots,\Gamma} \hat{e}(n)$

3.4. *Classifier and performance evaluation*

An SVM algorithm was used to develop a classification model to distinguish AD patients from HCs. During training, SVM seeks an optimal class separating a hyper-plane in the maximal margin [7,62–72]. In this work, SVM was performed using LIBSVM[4] with a linear kernel. The SVM with liner kernel has a regularization parameter that needs to be tuned during training. In order to tune this parameter, grid search process using inner 10-fold cross-validation during training process was used. In the grid search, the value of this parameter varied logarithmically from 2^{-5} to 2^{20}. To achieve a reliable measurement, all the performance results were obtained using 10-fold cross validation.

The classification results were assessed in terms of their ACC, SEN, SPE, and AUC based on 10-fold cross-validation. These parameters are defined as follows [18]:

$$ACC = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{17}$$

$$SEN = \frac{TP}{TP + FN} \tag{18}$$

$$SPE = \frac{TN}{TN + FP} \tag{19}$$

[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

where *TP* (the number of cases of AD correctly identified as AD), *TN* (the number of HCs correctly identified as HCs), *FN* (the number of AD patients incorrectly identified as HCs), and *FP* (the number of HCs incorrectly identified as having AD) denote the number of cases of true positive, true negative, false negative, and false positive, respectively [10]. AUC is the area under curve in which the curve stands for receiver operating characteristic (ROC). In the classification, ROC illustrates the performance of classifier by plotting the rate of sensitivity against (1 – specificity). The AUC provides a single scalar value of ROC in the range of [0, 1]. The AUC is a widely used measure of performance for classification [73].

3.5. *Data fusion among different feature ranking methods*

This paper introduces a data fusion technique among different feature ranking methods to improve the performance of the proposed feature-ranking-based AD classification. The aim of the data fusion technique is to integrate the data from two or more distinct multiple sources to improve performance. The pipeline of the proposed data fusion system combining different feature ranking methods is illustrated in Fig. 3. In the scheme of proposed data fusion, the top Γ-ranked features (i.e., $\Gamma = 1500$) $\left[f_*^1, f_*^2, \ldots, f_*^q\right]$ selected based on approaches, described in section 3.3.3, from different feature ranking methods, were combined into a single feature vector using union operator. Assuming $FRV_1, FRV_2, \ldots, FRV_z$ are feature ranked vectors generated using different feature ranking methods, the feature vector fusion (FVF) is then:

$$FVF = [FRV_1 \cup FRV_2 \cup \ldots \cup FRV_z]_{1\times\theta} \tag{20}$$

Where θ is the vector length for FVF, z is the number of ranking methods and $\Gamma \le \theta$. This concatenated feature vector was then used for post-feature ranking. In this regard, the MI based feature ranking was used, because of its better performance in comparison to other ranking methods. The ranked feature vector fusion is followed by the determination of resubstitution and cross-validation error estimators to select the top features that minimize the error in the all ranked feature vector fusion set.

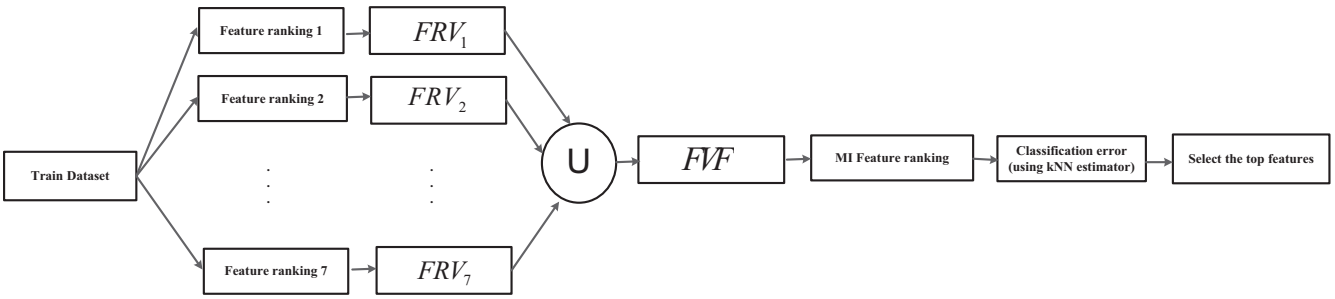

Fig. 3 – The pipeline of the proposed data fusion system combining different feature ranking methods.

4. Experimental results and discussion

In this section, the experimental results obtained through the preprocessing phase using VBM plus DARTEL analysis on 3D T1weighted MR Imaging are considered, as an indicator disclosing significance of decreased gray matter volumes in ADs contributing to VOI. The experimental data consisted of 260 samples from an ADNI data set. A 10-fold cross-validation was employed throughout the performance analysis, with 234 (90%) samples in the training sample and 26 (10%) samples in the testing processes in each iteration. The performance of the classification is reported for the following cases: (1) the performance of raw feature vectors directly extracted from atrophy regions using the VBM; (2) the performance of the proposed feature-ranking technique using the optimal number of top features based on the classification error; and (3) the performance of the proposed data fusion technique among different feature ranking methods. The ACC (%), SEN (%), SPE (%), and AUC performance metrics were used for the performance assessment.

4.1. *VBM of GM analysis in AD versus HC*

Generally, VBM plus DARTEL of GM analysis specified significant GM atrophy in the right/left hippocampus, right inferior parietal lobe, and right anterior cingulate in the ADs compared to the HCs through 10-fold cross validation. For an example, comparison of gray matter volume among 117 ADs and 117 HCs in fold 1 training is illustrated in Fig. 4. The voxel locations of these significant regions were segmented as a 3D mask in each fold. This mask was employed to the gray matter density volume results from the segmentation step in the MRI data pre-processing to extract voxel values from atrophy regions as raw feature vectors.

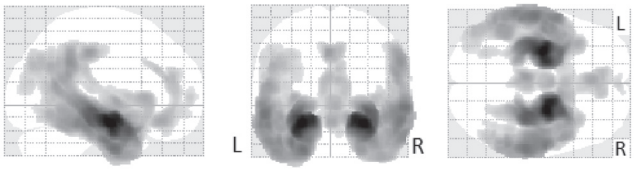

Fig. 4 – Brain regions with significant atrophy in gray matter volume in the 117 ADs compared to 117 HCs in fold 1 [25].

4.2. *Performance of raw feature vectors*

The complete MRI data set consisted of 130 AD and 130 HC samples. The ACC, SEN, SPE, and AUC obtained in the 10-fold cross validation using a linear SVM classifier on raw feature vectors are presented in Table 1.

4.3. *Performance of the proposed feature-selection method using feature ranking and classification error*

As introduced in Section 3.3.3, the proposed feature-selection technique was evaluated by using seven different feature-ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI), followed by two different classification errors (resubstitution and cross-validation error) to determine the optimal number of top features. The accuracy of the proposed feature-selection technique depends on the size of k in k-NN estimator. In order to determine the appropriate k value, a group of experiments were performed. Fig. 5 shows the accuracy rates of the different feature-ranking methods followed by two different classification errors with k = 1–10. Each point represents the mean of 10-fold cross validation. As Fig. 5 shows, the maximum accuracy is yielded with k = 3 for the most feature-ranking methods followed by two classification errors, and thus, k-NN estimator with k = 3 was employed to compute the classification error estimation. All of the scores for seven different feature-ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI) in the training set from fold 1 are plotted in Fig. 6(a–g). Fig. 7 shows the improvement in the ACC obtained by using progressive inclusion of the ranked features in the feature vectors from fold 1. A logarithmic scale was used to cover the entire feature space. This performance is reported for fold 1 after the MI feature ranking. The ACC was 80.76% and 92.30% on raw feature vectors and top 1500 ranked features after the MI feature ranking. The ACC performance improved with an increased number of ranked features, up to 96.15%. The performance level corresponded to the number of top ranked features, 479, which

Table 1 – Raw feature vectors performance of atrophy clusters using 10 fold cross validation.

ACC (%)	SEN (%)	SPE (%)	AU C
83.58	82.04	85.12	0.921

Note: ACC, accuracy; SEN, sensitivity; SPE, specificity; AUC, area under curve.

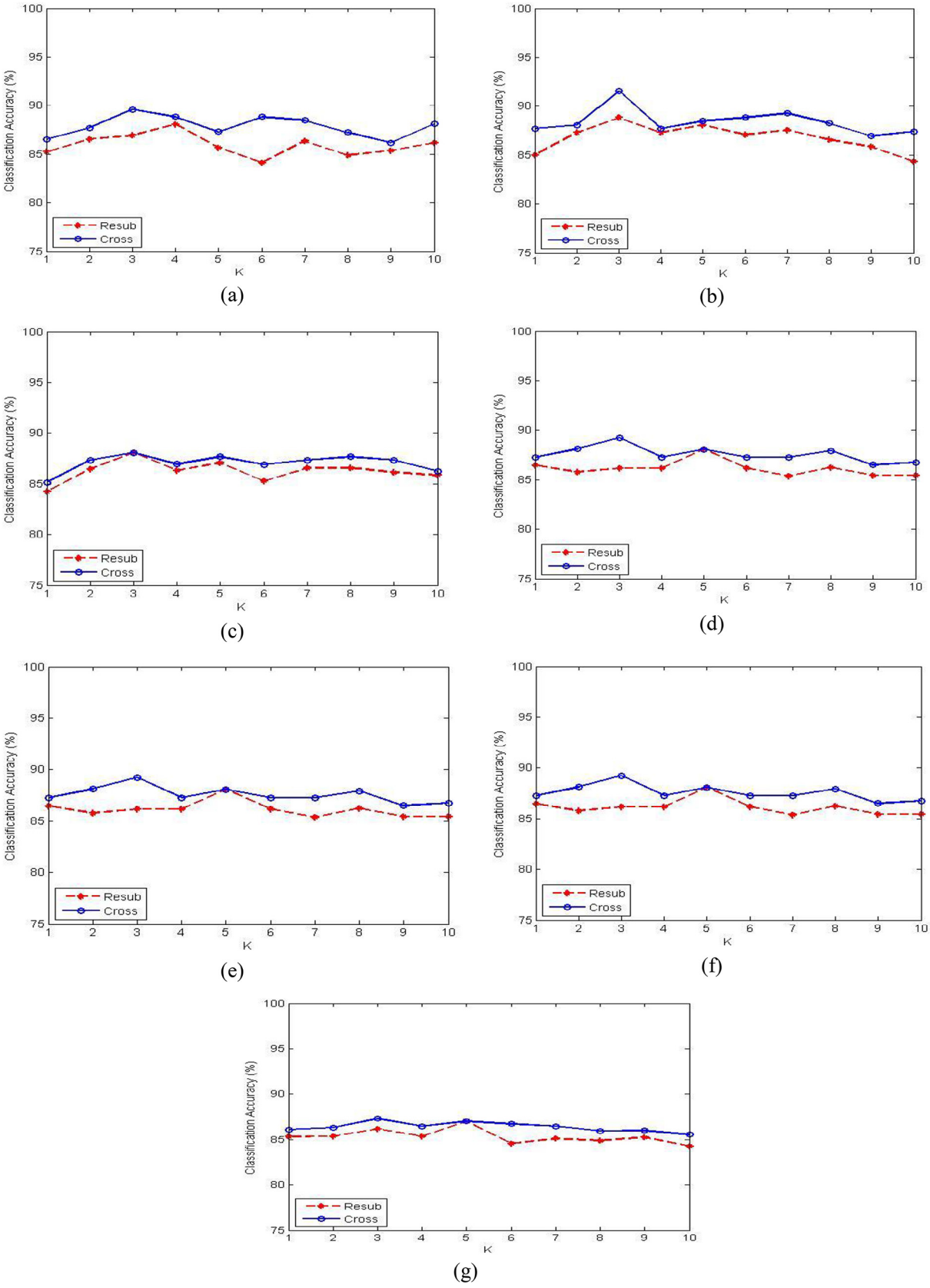

Fig. 5 – Classification accuracy of the seven feature-ranking methods, followed by two classification errors with different k. (a) SD, (b) MI, (c) IG, (d) PCC, (e) TS, (f) FC and (g) GI.

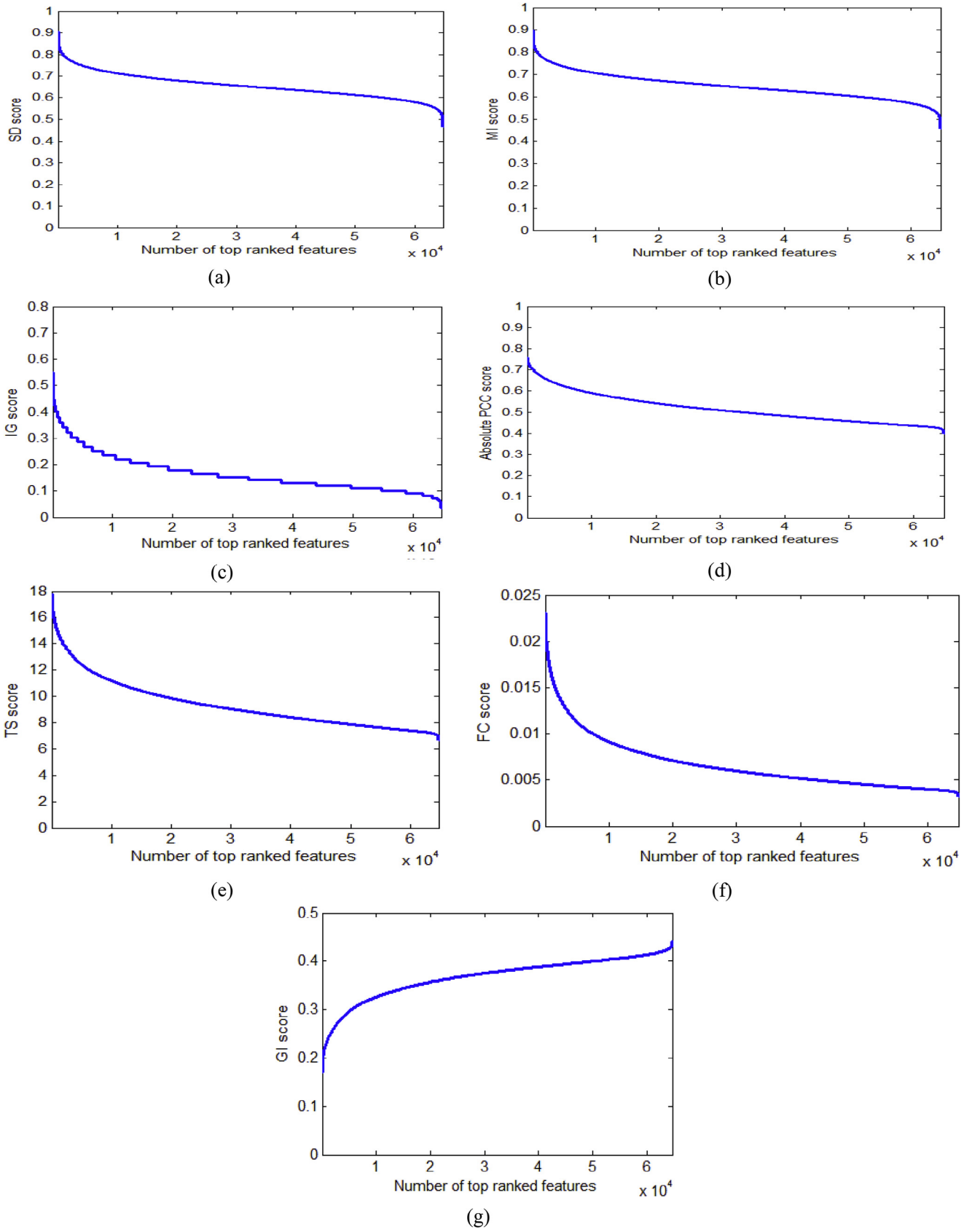

Fig. 6 – The scores for the respective ranked features with the seven different feature-ranking methods in the fold 1 training. (a) SD, (b) MI, (c) IG, (d) PCC, (e) TS, (f) FC and (g) GI.

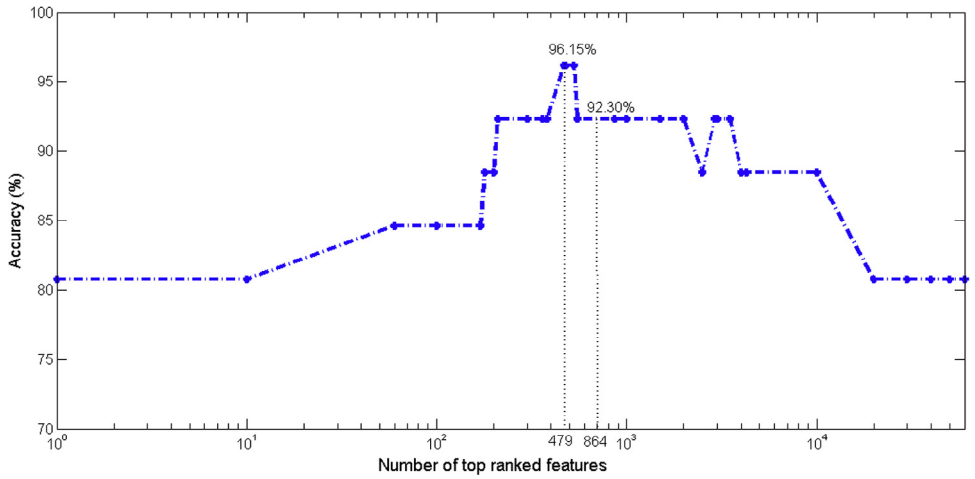

Fig. 7 – Accuracy (%) by different numbers of top ranked features selected using MI ranking in fold 1.

minimized the cross-validation error. The number of features that minimized the resubstitution error was 864, with an ACC performance of 92.30%. Table 2 presents the optimal number of top ranked features with different feature-ranking methods followed by two different classification errors in the training set from fold 1. Generally, the most selected top ranked features belong the to right and left hippocampus regions. Table 3 shows the overall performances of the proposed feature-selection method. To evaluate the overall performance of the different feature-ranking methods, followed by two different classification errors, the ROC curve of the classifiers are shown in Fig. 8(a–g). To draw the ROC Curves, 260 data points (i.e., 260 subjects) were used. The threshold value was set to zero for SVM classifier. The results clearly show the performance improvement provided by the proposed feature-selection method. Among the seven different feature-ranking methods, in general, the MI generated the highest performance for both classification errors to determine the optimal size of the feature vectors. Regarding the classification errors, the e_{cross}-based approach gives a higher performance than the e_{resub}-based method. The superior performance of the e_{cross}-based approach is attributed to the randomization in the cross validation, with the e_{cross}-based approach reducing the bias, which is the main problem of the e_{resub}-based method. Among the alternative methods tested, the results indicate that the MI feature ranking gives the highest or equal performance in terms of the ACC (%), SEN (%), SPE (%), and AUC, when compared with the other seven ranking methods.

Table 2 – The optimal number of top ranked features of the proposed feature selection method in the training set from fold 1.

	Resubstitution error (e_{resub})	Cross validation error (e_{cross})
SD	365	10
MI	864	479
IG	121	1269
PCC	1151	1286
TS	1151	1286
FC	1151	1286
GI	30	731

Note: SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Pearson's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index.

Table 3 – Performance results of the proposed feature selection method.

	Resubstitution error (e_{resub})				Cross validation error (e_{cross})			
	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
SD	86.92	83.07	**90.76**	0.943	89.61	88.46	90.76	0.957
MI	**88.84**	86.92	**90.76**	0.942	**91.53**	90.00	**93.07**	**0.958**
IG	88.07	**87.69**	88.46	**0.949**	88.07	86.92	89.23	0.945
PCC	86.15	86.92	85.38	**0.949**	89.23	**91.53**	86.92	0.946
TS	86.15	86.92	85.38	**0.949**	89.23	**91.53**	86.92	0.946
FC	86.15	86.92	85.38	**0.949**	89.23	**91.53**	86.92	0.946
GI	86.15	85.38	86.92	0.936	87.30	86.14	88.45	0.939

Note: For each column, the maximum value is specified in bold. ACC, accuracy; SEN, sensitivity; SPE, specificity; AUC, area under curve; SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Pearson's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index.

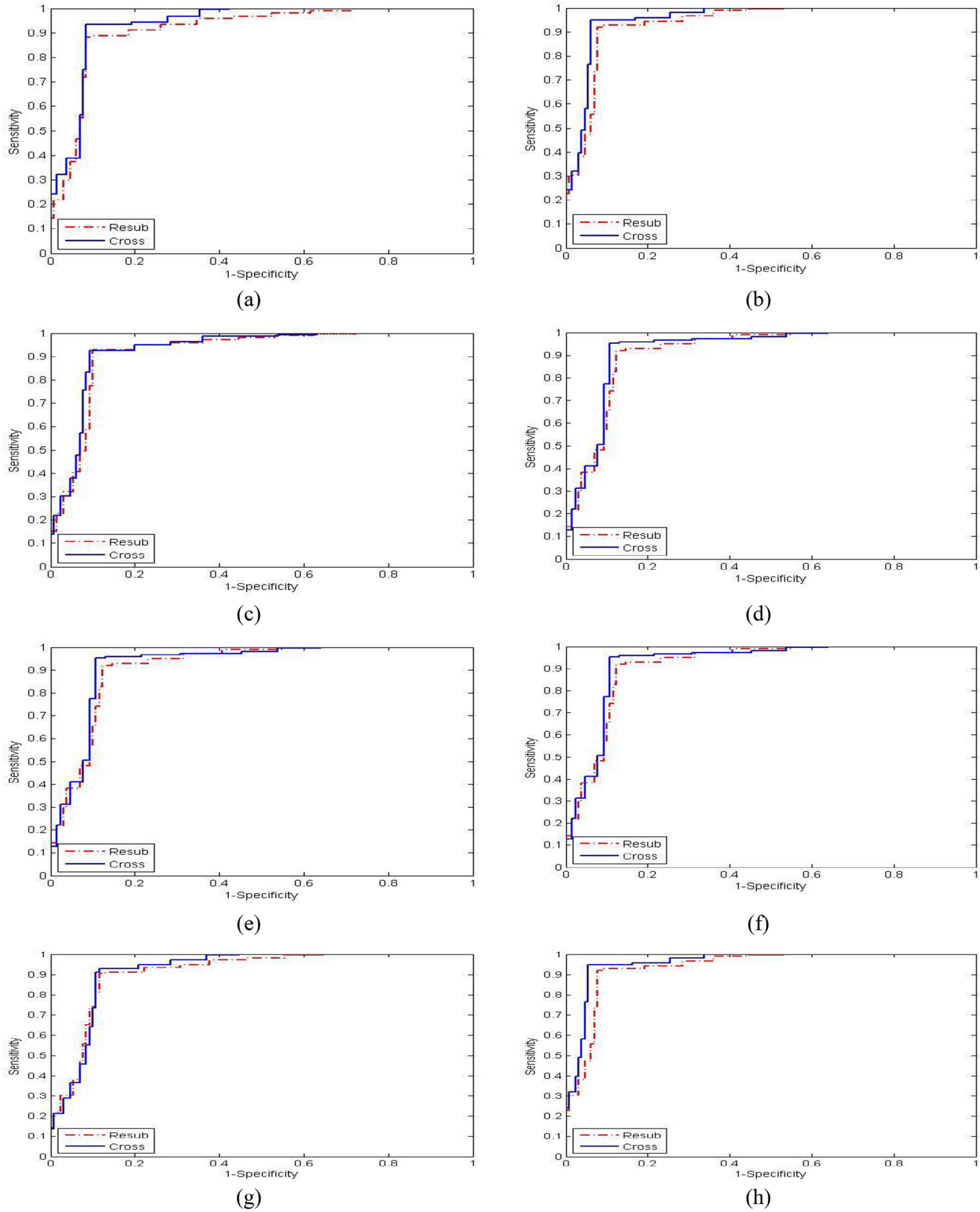

Fig. 8 – ROC curve of different feature-ranking methods, followed by two different classification errors. (a) SD, (b) MI, (c) IG, (d) PCC, (e) TS, (f) FC, (g) GI, and (h) proposed raking fusion.

Table 4 – Performance of proposed data fusion technique among feature ranking methods.

Resubstitution error (e_{resub})				Cross validation error (e_{cross})			
ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
88.84	86.92	90.76	0.942	**92.48**	**91.07**	**93.89**	**0.963**

Note: For each column, the maximum value is specified in bold. ACC, accuracy; SEN, sensitivity; SPE, specificity; AUC, area under curve.

4.4. *Performance of proposed data fusion among different feature ranking methods*

As proposed in section 3.5, the top Γ-ranked features (i.e., $\Gamma = 1500$) from different statistical feature ranking methods were combined into a single feature vector using a union operator. For example, the length of this concatenated feature vector was 2584 in fold 1 training. This concatenated feature vector was then used for post-feature ranking using MI-based feature ranking method, because of its better performance in comparison to other ranking methods (see Table 3). The ranked feature vector fusion was followed by the determination of resubstitution and cross-validation error estimators to select the top features that minimize the error in the all ranked feature vector fusion set. The number of top ranked features that minimized the resubstitution and cross-validation errors were 864 and 1920, respectively, in fold 1 training. The performance improvement aided by proposed data fusion of seven different feature ranking methods is shown in Table 4. The performance of the e_{cross}-based approach is always higher than the performance of the e_{resub}-based method. Fig. 8(h) shows the ROC curve related to proposed ranking fusion method.

4.5. *Discussion*

The present study investigated the feature ranking and classification errors as part of a novel feature-selection method to design an automatic CAD system for high-dimensional pattern classification in AD. In the proposed system, we evaluated seven feature ranking approaches to rank the features with respect to their statistical relevance. In addition, we proposed an automatic criterion to select the subset of top ranked features based on classification error in the training part. In this context, resubstitution and cross-validation error estimators were employed to identify the number of ranked features. By investigating Table 1 and Table 3, it is clear that the proposed feature selection method significantly improved the performance with respect to raw feature vectors. For example, feature selection using MI ranking and cross-validation error estimator provided 8% improvement in accuracy in comparison to raw feature vectors. Many researchers studied Random Forest as an alternative feature selection method in machine learning, because of its relatively good accuracy and robustness [74–76]. Otherwise, using Random Forest suffers from bias toward features with many categories and with correlated features, more informative features can end up with low scores [77]. In addition, several studies investigated high-dimensional pattern classification approach in a number of neuroimaging studies [25,78–80]. For example in Ref. [80], the authors presented an advanced quantitative pattern analysis and classification of brain atrophy in MCI and AD patients. In Ref. [79] the authors introduced a method based on Support Vector Machine-Recursive Feature Elimination (SVM-RFE) technique for feature ranking and they used SVM classifier for classification. Data used in the present study were the same as the one described in our previous study [25], including pre-processing steps and feature extraction. In Ref. [25], we introduced a novel statistical feature selection method based on the probability distribution function (PDF) of the VOI. In more detail, PDF was introduced to generate statistical pattern of the VOI representing the entire sMRI. Using the proposed PDF-based method, we obtained 89.65% accuracy with linear SVM. In the present study, instead of generation of the statistical pattern of the VOI, we introduced an automatic statistical feature selection method based on the feature ranking and the classification error of the VOI, which can be considered a lower-dimensional feature vector representation of sMRI. The dimensionality of the feature vector can be adjusted by minimizing the classification error in the training data-set. The proposed feature selection method not only selects the top discriminative features but also reduces the dimensionality of the input vectors to feature vectors. In one of our previous studies [81], we introduced a feature selection approach based on t-test feature ranking and Fisher Criterion (FC) for high-dimensional pattern recognition in AD detection. In Ref. [81], the number of top features was determined by using FC, which maximizes the class separation between AD and HC. In this study, instead of using FC, we propose to use classification error as stopping criterion to determine the optimal number of top ranked features. We use the method which is presented in Ref. [81] on the current dataset. Table 5 presents a comparison of the classification results based on seven feature-ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI) and FC as stopping. As shown in Tables 3 and 5, the ACC performance of MI feature ranking is higher than the other ranking methods for all of the three different stopping criteria. Recently, MI feature selection approach has been widely used for feature selection in pattern recognition

Table 5 – Comparison of classification performance using seven different feature ranking methods and FC as stopping criterion.

Method	Stopping criteria	ACC (%)	SEN (%)	SPE (%)	AUC
SD	FC	87.30	**86.92**	87.69	0.952
MI	**FC**	**88.07**	**86.92**	89.23	**0.953**
IG	FC	86.53	83.07	**90.00**	0.936
PCC	FC	86.92	85.38	88.46	0.940
TS	FC	86.92	85.38	88.46	0.940
FC	FC	86.92	85.38	88.46	0.940
GI	FC	86.92	85.38	88.46	0.940

Note: For each column, the maximum value is specified in bold. ACC, accuracy; SEN, sensitivity; SPE, specificity; AUC, area under curve; SD, statistical dependency; MI, mutual information; IG, information gain; PCC, Pearson's correlation coefficient; TS, t-test score; FC, Fisher criterion; GI, Gini index; e_{resub}, resubstitution error; e_{cross}, cross validation error.

Table 6 – Training accuracy base on MI feature ranking and three different stopping criteria.

Stopping criteria	ACC (%)	SEN (%)	SPE (%)	AUC
FC	90.29	88.71	91.88	0.957
e_{resub}	92.05	90.68	**93.41**	0.962
e_{cross}	**92.43**	**91.53**	93.33	**0.968**

Note: For each column, the maximum value is specified in bold. ACC, accuracy; SEN, sensitivity; SPE, specificity; AUC, area under curve; SD, statistical dependency; FC, Fisher criterion; e_{resub}, resubstitution error; e_{cross}, cross validation error.

studies [82,83]. In addition, regarding the stopping criteria, e_{cross} shows better performance in comparison to the other stopping criteria such as e_{resub} and FC among the alternative ranking methods. Regarding the stopping criteria, classification error is based not only to select the most discriminative/informative features, but also to minimize the training error. Minimizing training error corresponds to maximizing training accuracy and maximum level of learning in the classification process. As an example, Table 6 shows the training accuracy based on MI-based feature ranking and three different stopping criteria using linear SVM. As shown in Table 6, it is clear that training accuracy of the e_{cross}-based approach is better than e_{resub} and FC.

Finally, we proposed a data fusion technique among the different feature ranking methods and obtained 92.48% accuracy with linear SVM. As part of future studies on AD classification, we suggest considering feature ranking-based feature selection for high-dimensional pattern classification such as the deformation-based analysis. Another priority for future studies is to use other registration methods as described in Ref. [84]. These methods could further be used to evaluate the accuracy of inter-subject registration in GM volume changes in patients with AD.

4.6. *Performance comparison to other methods*

Recently, several studies have reported classification results to distinguish AD patients and HCs based on MRI and ADNI dataset. Aguilar et al. [5] employed the FreeSurfer software to compute cortical thickness and volumetric measures. Based on an artificial neural network classifier and MRI data, they achieved an ACC of 84.9% and an ACC of 88.8% using an SVM classifier and a combination of MRI data with educational and demographic data. Querbes et al. [85] reported an ACC of 85% using a cortical thickness feature from MRI data. Khedner et al. [10] achieved an ACC of 88.49% by combining GM and white matter modalities in MRI data. Cuingnet et al. [86] tested 10 methods. They presented an SEN of 81% and an SPE of 95% as the best performances. Zhang et al. [3] used a multimodal classification of AD based on a combination of MRI, CSF, and PET data. They reported an ACC of 86.2% in the classification of AD/HC using the MRI data. By combining the MRI, CSF, and PET results, they achieved a high ACC of 93.2%. Westman et al. [6] reported an ACC of 87% using MRI data and increased the ACC to 91.8% by combining the MRI data with CSF measures. Beheshti and Demirel [25] employed a PDF-based approach using MRI data and reported an ACC of 89.65%. A comparison of the classification performance using the different methods and MRI data is provided in Table 7. The results show that the performance of the proposed feature-selection method using only MRI data is higher or well comparable to that of other methods reported in the literature.

5. Conclusion

This paper proposed an automatic CAD system for the classification of AD based on seven feature-ranking methods (i.e., SD, MI, IG, PCC, TS, FC, and GI) and classification errors (i.e., resubstitution and cross-validation errors). The optimal size of the selected features was determined by classification error estimation, which minimized the classification error in the training phase. This approach was applied to extracted raw features obtained from GM atrophy clusters of VOIs, which were determined using a VBM analysis. An SVM classifier was used for the classification of the extracted feature vectors after the feature selection. A performance improvement was also proposed by applying data fusion among the different feature ranking basedmethods. The performance of the proposed system was evaluated with 10-fold cross validation using an ADNI data set made up of 260 subjects (130 AD patients and 130 HCs). The results clearly showed that the proposed feature-selection method was a reliable technique for high-dimensional data. The experimental results showed that the performance of the proposed approach using only MRI data was higher or comparable to that of alternative methods reported in the literature.

Table 7 – Supervised classification results of Alzheimer's disease and healthy control subjects based on MRI from ADNI data-set.

Author	Imaging modality	AD/HC	Validation method	ACC (%)	SEN (%)	SPE (%)	AUC
Aguilar et al. [5]	MRI	116/110	10 Fold	84.90	80.20	90.00	0.880
Querbes et al. [85]	MRI	130/130	10 Fold	85.00	–	–	–
Khedher et al. [10]	MRI	188/229	10 Fold	88.49	85.11	91.27	–
Cuingnet et al. [86][a]	MRI	162/137	2 Fold	–	81.00	95.00	–
Zhang et al. [3]	MRI	51/52	10 Fold	86.20	86.00	86.30	-
Westman et al. [6]	MRI	96/111	10 Fold	87.00	83.30	90.10	0.930
Beheshti and Demirel [25]	MRI	130/130	10 Fold	89.65	87.73	91.57	0.953
Proposed method	**MRI**	**130/130**	**10 Fold**	**92.48**	**91.07**	**93.89**	**0.963**

[a] This paper compares ten methods and the best performance is presented here.

Acknowledgments

This work was partly carried out under the Brain Mapping by Integrated Neuroethologies for Disease Studies (Brain/MINDS) project (grant number 16dm0207017h0003), funded by the Japan Agency for Medical Research and Development (AMED). This work has been also partially supported by project grants from Beijing Nova Program (xx2016120), National Natural Science Foundation of China (81101107, 31640035), Natural Science Foundation of Beijing (4162008) and program for top young innovative talents of the Beijing Educational Committee (CIT&TCD201404053).

Compliance with ethical standards

Conflict of interest

All authors declared no conflict of interests.

Ethical approval

All authors confirm that the data used in our research involving human participants were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI).

REFERENCES

[1] M.W. Weiner, D.P. Veitch, P.S. Aisen, L.A. Beckett, N.J. Cairns, R.C. Green, et al., 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception, Alzheimer's Dement 11 (2015) e1–e120, http://dx.doi.org/10.1016/j.jalz.2014.11.001.

[2] Alzheimer's Association. Alzheimer's disease and dementia. http://www.alz.org/, 2015 (accessed 05.04.15).

[3] 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.

[4] 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.

[5] 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. 212 (2013) 89–98, http://dx.doi.org/10.1016/j.pscychresns.2012.11.005.

[6] 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.

[7] 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.

[8] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, 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.

[9] E.E. Bron, M. Smits, W.M. van der Flier, H. Vrenken, F. Barkhof, P. Scheltens, et al., 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.

[10] 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.

[11] D. Chyzhyk, A. Savio, M. Graña, Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI, Neurocomputing 128 (2014) 73–80.

[12] D. Chyzhyk, M. Graña, A. Savio, J. Maiora, Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI, Neurocomputing 75 (2012) 72–77.

[13] 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.

[14] 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, http://dx.doi.org/10.1016/j.neuroimage.2008.02.043.

[15] E. Dinesh, M.S. Kumar, M. Vigneshwar, T. Mohanraj, Instinctive classification of Alzheimer's disease using FMRI, pet and SPECT images, 2013 7th Int. Conf. Intell. Syst. Control., 405–409. http://dx.doi.org/10.1109/ISCO.2013.6481189, 2013. (accessed 11.18.15).

[16] L. Mesrob, DTI and structural MRI classification in Alzheimer's disease, Adv. Mol. Imaging 02 (2012) 12–20, http://dx.doi.org/10.4236/ami.2012.22003.

[17] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, et al., 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.

[18] 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.

[19] 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.

[20] K.R. Gray, R. Wolz, R.A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, 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.

[21] Y.J. Chen, G. Deutsch, R. Satya, H.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.

[22] J.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.

[23] 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.
[24] 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.
[25] 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.
[26] A. Ortiz, J.M. Górriz, J. Ramírez, F.J. Martínez-Murcia, LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease, Pattern Recognit. Lett. 34 (2013) 1725–1733, http://dx.doi.org/10.1016/j.patrec.2013.04.014.
[27] 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.
[28] 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) 2322, e19-27, http://dx.doi.org/10.1016/j.neurobiolaging.2010.05.023.
[29] 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.
[30] 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.
[31] J. Ashburner, K.J. Friston, Voxel-based morphometry – the methods, Neuroimage 11 (2000) 805–821, http://dx.doi.org/10.1006/nimg.2000.0582.
[32] 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.
[33] 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, AJNR Am. J. Neuroradiol. 33 (2012) 1109–1114, http://dx.doi.org/10.3174/ajnr.A2935.
[34] X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, et al., 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.
[35] Y. Meng, C. Qiu, H. Zhu, S. Lama, S. Lui, Q. Gong, et al., Anatomical deficits in adult posttraumatic stress disorder: a meta-analysis of voxel-based morphometry studies, Behav. Brain Res. 270 (2014) 307–315, http://dx.doi.org/10.1016/j.bbr.2014.05.021.
[36] I. Beheshti, H. Demirel, C. Yang, Significance of sex differences on gray matter atrophy in Alzheimer's disease: a voxel-based morphometry study, Br. Biomed. Bull. 3 (2015) 522–536.
[37] Y.-W. Chang, C.-J. Lin, Feature ranking using linear svm, JMLR Workshop Conf. Proc. 3 (2008) 53–64.
[38] X. Geng, T. Liu, T. Qin, H. Li, Feature selection for ranking, Sigir (2007) 407–414.
[39] R.C. Prati, Combining feature ranking algorithms through rank aggregation, Proc. Int. Jt. Conf. Neural Networks. 10–15. http://dx.doi.org/10.1109/IJCNN.2012.6252467, 2012 (accessed 10.05.15).
[40] 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.
[41] W. Duch, T. Wieczorek, J. Biesiada, M. Blachnik, Comparison of feature ranking methods based on information entropy, IEEE Int. Conf. Neural Networks Conf. Proc. 2 (2004) 1415–1419, http://dx.doi.org/10.1109/IJCNN.2004.1380157.
[42] 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.
[43] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. Van Der Kouwe, R. Gollub, et al., 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.
[44] J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1998) 87–97.
[45] 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.
[46] J.H. Son, D.H. Han, K.J. Min, B.S. Kee, Correlation between gray matter volume in the temporal lobe and depressive symptoms in patients with Alzheimer's disease, Neurosci. Lett. 548 (2013) 15–20, http://dx.doi.org/10.1016/j.neulet.2013.05.021.
[47] L. Xu, X. Wu, K. Chen, L. Yao, Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment, Comput. Methods Programs Biomed. (2015) 1–9, http://dx.doi.org/10.1016/j.cmpb.2015.08.004.
[48] 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.
[49] D.N. Greve, An absolute beginner's guide to surface- and voxel-based morphometric analysis, Present. ISMRM 19th Annu. Meet. Exhib. Montréal, Québec, Canada. 7–13, 2011.
[50] D. Chyzhyk, A. Savio, M. Graña, Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM, Neural Netw. 68 (2015) 23–33.
[51] 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.
[52] W. Yan, Fusion in multi-criterion feature ranking, Inf. Fusion, 2007 10th Int. Conf. 1–6. http://dx.doi.org/10.1109/ICIF.2007.4408064, 2007 (accessed 01.05.16).
[53] N. Zhou, L. Wang, A modified T-test feature selection method and its application on the HapMap genotype data, Genomics Proteomics Bioinformatics 5 (2007) 242–249, http://dx.doi.org/10.1016/S1672-0229(08)60011-X.
[54] W. Wang, X. Wang, D. Feng, J. Liu, Z. Han, X. Zhang, Exploring permission-induced risk in android applications for malicious application detection, IEEE Trans. Inform. 9 (2014) 1869–1882.
[55] 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.compbiomed.2015.01.003.
[56] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, Advancing feature selection research, ASU Featur. Sel. Repos. Arizona State Univ. 1–28. http://featureselection.asu.edu/featureselection_techreport.pdf, 2010 (accessed 12.27.15).
[57] I. Kamkar, S.K. Gupta, D. Phung, S. Venkatesh, Stable feature selection for clinical prediction: exploiting ICD tree structure using Tree-Lasso, J. Biomed. Inform. 53 (2014) 277–290, http://dx.doi.org/10.1016/j.jbi.2014.11.013.
[58] 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.
[59] C. Sima, U.M. Braga-Neto, E.R. Dougherty, High-dimensional bolstered error estimation, Bioinformatics 27 (2011) 3056–3064, http://dx.doi.org/10.1093/bioinformatics/btr518.
[60] C. Sima, U. Braga-Neto, E.R. Dougherty, Superior feature-set ranking for small samples using bolstered error estimation, Bioinformatics 21 (2005) 1046–1054, http://dx.doi.org/10.1093/bioinformatics/bti081.
[61] U.M. Braga-Neto, Classification and error estimation for discrete data, Curr. Genomics 10 (2009) 446–462, http://dx.doi.org/10.2174/138920209789208228.
[62] I. Dimitrovski, D. Kocev, I. Kitanovski, S. Loskovska, S. Džeroski, Improved medical image modality classification using a combination of visual and textual features, Comput. Med. Imaging Graph. 39 (2015) 14–26, http://dx.doi.org/10.1016/j.compmedimag.2014.06.005.
[63] O.S. Al-Kadi, A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours, Comput. Med. Imaging Graph. 41 (2014) 67–79, http://dx.doi.org/10.1016/j.compmedimag.2014.05.013.
[64] T. Xue, L. Bai, S. Chen, C. Zhong, Y. Feng, H. Wang, et al., Neural specificity of acupuncture stimulation from support vector machine classification analysis, Magn. Reson. Imaging 29 (2011) 943–950, http://dx.doi.org/10.1016/j.mri.2011.03.003.
[65] 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.
[66] C. Hinrichs, V. Singh, G. Xu, S.C. Johnson, Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population, Neuroimage 55 (2011) 574–589, http://dx.doi.org/10.1016/j.neuroimage.2010.10.081.
[67] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1988.
[68] A. Farzan, S. Mashohor, A.R. Ramli, R. Mahmud, Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns, Behav. Brain Res. 290 (2015) 124–130, http://dx.doi.org/10.1016/j.bbr.2015.04.010.
[69] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease, Brain Imaging Behav. 10 (2015) 799–817, http://dx.doi.org/10.1007/s11682-015-9448-7.
[70] L. Wang, C.-Y. Wee, X. Tang, P.-T. Yap, D. Shen, Multi-task feature selection via supervised canonical graph matching for diagnosis of autism spectrum disorder, Brain Imaging Behav. 10 (2015) 33–40, http://dx.doi.org/10.1007/s11682-015-9360-1.
[71] A. Besga, D. Chyzhyk, I. González-Ortega, A. Savio, B. Ayerdi, J. Echeveste, et al., Eigenanatomy on fractional anisotropy imaging provides white matter anatomical features discriminating between Alzheimer's disease and late onset bipolar disorder, Curr. Alzheimer Res. 13 (2016) 557–565.
[72] D. Chyzhyk, M. Graña, D. Öngür, A.K. Shinn, Discrimination of schizophrenia auditory hallucinators by machine learning of resting-state functional mri, Int. J. Neural Syst. 25 (2015) 1550007.
[73] D.J. Hand, Measuring classifier performance: a coherent alternative to the area under the ROC curve, Mach. Learn. 77 (2009) 103–123, http://dx.doi.org/10.1007/s10994-009-5119-5.
[74] R. Genuer, J. Poggi, C. Tuleau-Malot, Variable selection using random forests, Pattern Recognit. Lett. 31 (2010) 2225–2236, http://dx.doi.org/10.1016/j.patrec.2010.03.014.
[75] T. Ebina, H. Toh, Y. Kuroda, DROP: an SVM domain linker predictor trained with optimal features selected by random forest, Bioinformatics 27 (2011) 487–494, http://dx.doi.org/10.1093/bioinformatics/btq700.
[76] R. Díaz-Uriarte, S.A. De Andrés, Gene selection and classification of microarray data using random forest, Bioinformatics 13 (2006) 1–13, http://dx.doi.org/10.1186/1471-2105-7-3.
[77] C. Strobl, A. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: illustrations, sources and a solution, BMC Bioinformatics 21 (2007) http://dx.doi.org/10.1186/1471-2105-8-25.
[78] 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.
[79] 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.
[80] 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.
[81] I. Beheshti, H. Demirel, Feature-ranking-based Alzheimer's disease classification from structural MRI, Magn. Reson. Imaging 34 (2015) 252–263, http://dx.doi.org/10.1016/j.mri.2015.11.009.
[82] Z. Yan, Z. Wang, H. Xie, The application of mutual information-based feature selection and fuzzy LS-SVM-based classifier in motion classification, Comput. Methods Programs Biomed. 90 (2008) 275–284, http://dx.doi.org/10.1016/j.cmpb.2008.01.003.
[83] S.-N. Yu, M.-Y. Lee, Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability, Comput. Methods Programs Biomed. 108 (2012) 299–309, http://dx.doi.org/10.1016/j.cmpb.2011.12.015.
[84] 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.
[85] 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.
[86] 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.