Progress In Electromagnetics Research, Vol. 144, 171–184, 2014

Classification of Alzheimer Disease Based on Structural Magnetic Resonance Imaging by Kernel Support Vector Machine Decision Tree

Yudong Zhang[1, *], Shuihua Wang[1, 2], and Zhengchao Dong[3]

Abstract—In this paper we proposed a novel classification system to distinguish among elderly subjects with Alzheimer's disease (AD), mild cognitive impairment (MCI), and normal controls (NC). The method employed the magnetic resonance imaging (MRI) data of 178 subjects consisting of 97 NCs, 57 MCIs, and 24 ADs. First, all these three dimensional (3D) MRI images were preprocessed with atlas-registered normalization. Then, gray matter images were extracted and the 3D images were under-sampled. Afterwards, principle component analysis was applied for feature extraction. In total, 20 principal components (PC) were extracted from 3D MRI data using singular value decomposition (SVD) algorithm, and 2 PCs were extracted from additional information (consisting of demographics, clinical examination, and derived anatomic volumes) using alternating least squares (ALS). On the basic of the 22 features, we constructed a kernel support vector machine decision tree (kSVM-DT). The error penalty parameter C and kernel parameter σ were determined by Particle Swarm Optimization (PSO). The weights ω and biases b were still obtained by quadratic programming method. 5-fold cross validation was employed to obtain the out-of-sample estimate. The results show that the proposed kSVM-DT achieves 80% classification accuracy, better than 74% of the method without kernel. Besides, the PSO exceeds the random selection method in choosing the parameters of the classifier. The computation time to predict a new patient is only 0.022 s.

1. INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia. There is no cure for this disease, which is progressively worsening and eventually leads to death [1]. In 2006, there were 26.6 million sufferers worldwide, and AD is predicted to affect 1 in 85 people globally by 2050, and at least 43% of prevalent cases need a high level of care [2]. As the world is evolving into an aging society, the burdens and impacts caused by AD on families and the society will be increasingly pronounced.

While there is currently no cure for AD, early, accurate and effective detection of AD is beneficial for the management of the disease. A multitude of neurologists and medical researchers have been dedicating much time and energy toward this goal, and promising results have been continually springing up [3]. Structural magnetic resonance imaging (sMRI) plays an important role in distinguishing AD subjects from normal controls (NC), and to distinguish mild cognitive impairment (MCI) subjects who later convert to AD from those who do not. In early days, much work was done to measure manually or semi-manually a priori region of interest (ROI) based on the fact that AD patients suffer much cerebral atrophy compared to NCs [4]. Most of such ROI-based analyses focuses on the regions of hippocampus and entorhinal cortex [5]. After comparing the ROIs between different subjects, the examiners can discover valuable information and provided auxiliary information for the diagnosis. However, the ROI-based methods suffer from some limitations. First, they focus on the ROIs pre-determined based on

Received 13 December 2013, Accepted 6 January 2014, Scheduled 16 January 2014

* Corresponding author: Yudong Zhang (zhangyudongnuaa@gmail.com).

[1] School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China. [2] School of Electronic Science and Engineering, Nanjing University, Nanjing, Jiangsu 210046, China. [3] Translational Imaging Division & MRI Unit, Columbia University and New York State Psychiatric Institute, New York, NY 10032, USA.

prior knowledge, rather than data-driven or exploratory; second, the accuracy of early detection depends heavily on the experience of the examiner; third the efficiency is low; forth, the mutual information among the voxels is difficult to operate [3, 6].

In contrast, the multivariate methods take into consideration of the relationship among voxels across the whole brain and, therefore, no specific ROIs are needed. The multivariate methods are commonly time-consuming and even more expertise-demanding, making manual analysis impossible [7]. Therefore, automated classification methods are desirable.

The idea of the automated classifier is described as follows: Principle Component Analysis (PCA) is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data [8]. Those principal components (PCs) extracted from MRI data are representative of the natural groupings of the brain circuits affected during the course of progression of AD. It is a supervised learning task for automatically classification among AD, NC and MCI. The support vector machines (SVM) are extremely powerful tools for supervised learning [9]. However, SVM is for solving two-class pattern recognition problem. In order to use SVM solving multi-class problems, we introduce in the Support Vector Machine Decision Tree (SVM-DT) method, and integrate it with the kernel technique for better performance. We call the new method as kernel SVM-DT (kSVM-DT). The weights ω and biases b of kSVM-DT are obtained by quadratic programming (QP). The error penalty parameter C and kernel parameter σ of kSVM-DT are obtained by Particle Swarm Optimization (PSO).

The structure of the paper is organized as follows: The next Section 2 divides the experimental data to training data (MRI data, demographics, clinical examination, and derived anatomic volume) and target data. Section 3 presents the preprocessing procedure. It shows the detailed principles and methods of PCA, and illustrates the application of PCA to the MRI data. Section 4 presents the methodology of kSVM-DT. First, the skewness tree is introduced. Then, the kernel technique is employed. Afterwards, the PSO is used to find the optimal error penalty parameter C and kernel parameter σ. Finally, the cross validation is introduced for out-of-sample estimate. Experiments in Section 5 show the results of the proposed method, and make in-depth analysis of each step. Section 6 concludes the paper, summarizes our contributions, and plans for future work.

2. EXPERIMENTAL DATA

We download the public dataset from Open Access Series of Imaging Studies (OASIS) [10, 11]. The url is http://www.oasis-brains.org/. OASIS contains cross-sectional MRI data, longitudinal MRI data, and additional data for each subject. In this study, we choose the cross-sectional dataset corresponding to the MRI scan of individuals at a single point in time [12]. The OASIS dataset consists of 416 subjects aged 18 to 96. We excluded subjects under 60 years old, and then picked up to 178 subjects from the rest subjects. The attributes of the included subjects are brief summarized in Table 1. The subjects of all classes are not equally since the OASIS contains little ADs and MCIs.

For each subject, 3 or 4 individual T1-weighted MRI images obtained in a single scan session are included. The subjects are all right-handed and include both men and women. 100 of the included subjects were clinically diagnosed with very mild to moderate Alzheimer's disease [13].

Table 1. Subject demographics and dementia status.

	NC	MCI	AD
Number of Subjects	97	57	24
Male/Female	26/71	26/31	7/17
Age	75.94 ± 9.03	77.07 ± 7.08	78.46 ± 6.56
Education	3.27 ± 1.32	3.05 ± 1.36	2.58 ± 1.38
SES	2.52 ± 1.09	2.67 ± 1.12	2.88 ± 1.30
CDR	0	0.5	1
MMSE	28.97 ± 1.21	25.88 ± 3.26	21.96 ± 3.26

2.1. MRI Data

For each individual, all available 3 or 4 MRI images were motion-corrected, and coregistered to form an averaged image. Then, those images were spatially normalized to the Talairach coordinate space and brain-masked. Finally, all images were segmented and smoothed. Each of the tissue types are weighted with different value, 0: background; 1: CSF (Cerebro-Spinal Fluid); 2: GM (Gray Matter); 3: WM (White Matter). The GM images were extracted since GM is highly correlated to AD [14, 15]. The brain extraction tool (BET) [16] was employed removing facial features (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET). Figure 1 shows the snapshots of our preprocessing procedure.

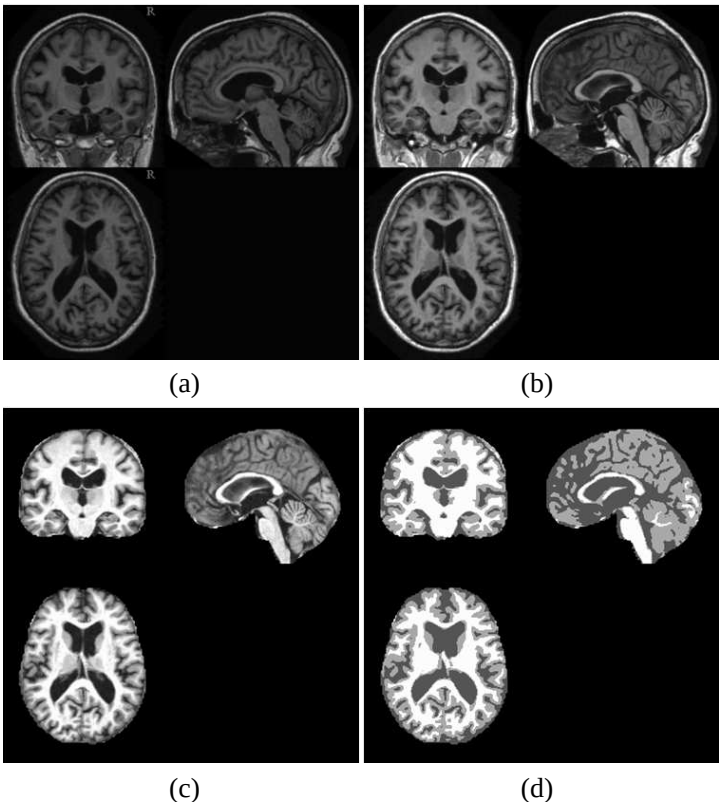

Figure 1. Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain-masked version of Figure 1(b). (d) The grey/white/CSF segmentation image.

2.2. Additional Data

We choose demographics, clinical examination, and derived anatomic volume as the additional data, so as to increase the performance of the classification.

2.2.1. Demographics

The demographical features contain gender (M/F), handedness, age, education, and socioeconomic status (SES). Education codes are listed in Table 2. The handedness feature is not employed in this paper since all subjects are right-handed.

2.2.2. Clinical Examination

The mini-mental state examination (MMSE), also known as Folstein test, is a brief 30-point questionnaire test used to screen for cognitive impairment and dementia [17]. The MMSE test includes

Table 2. Education codes.

Code	Description
1	Less than high school graduate
2	High school graduate
3	Some college
4	College graduate
5	Beyond college

Table 3. A typical MMSE test.

Category	Points	Description
Orientation to Time	5	From broadest to most narrow
Orientation to Place	5	From broadest to most narrow
Registration	3	Repeating named prompts
Attention and Calculation	5	Serial Sevens, or spelling "world" backwards
Recall	3	Registration Recall
Language	2	Naming a pencil and a watch
Repetition	1	Speaking back a phrase
Complex Commands	6	Drawing figures shown, etc.

simple questions and problems in a number of areas: the time and place, repeating lists of words, arithmetic, language use & comprehension, and basic motor skills. A typical MMSE test is shown in Table 3.

2.2.3. Derived Anatomic Volumes

Three features were extracted, including the estimated total intracranial volume (eTIV) (mm^3), atlas scaling factor (ASF), and normalized whole brain volume (nWBV). The eTIV was used as the correction for 'intracranial volume (ICV)' because certain structures scale with general head size [18]. For example, people with larger heads typically have larger hippocampi. However, researchers are especially interested in the deviation of the volume of the structure from what may be expected for the size of that structure, as it may potentially reveal the disease-related changes. The expected value can be based on the individual's intracranial value and the scaling factor for that particular structure [19]. The ASF is defined as the volume-scaling factor required matching each individual to the atlas target [19]. The nWBV is acquired using the methods in literature [20].

In total, the additional data contains the following features: gender, age, education, SES, MMSE, eTIV, ASF, and nWBV.

2.3. Target Data

The clinical dementia rating (CDR) was used as the target during training process. It is a numeric scale quantifying the severity of symptoms of dementia [21]. The patient's cognitive and functional performances were assessed in six areas: memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care.

In this study, we choose three types of CDR: 1) subjects with CDR as 0 are considered NC; 2) subjects with CDR as 0.5 are considered MCI; 3) subjects with CDR as 1 are considered AD [22].

3. PREPROCESSING BY PCA

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. In the present paper, we used PCA as a preprocessing step to reduce the number of features.

3.1. Principles

PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables, while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors so that they are not correlated with each other, it orders the resulting orthogonal components so that those with the largest variation come first, and it eliminates those components contributing the least to the variation in the data set [23]. It should be noted that the input vectors should be normalized to have zero mean and unity variance before performing PCA.

Given n observations $\{x_1, x_2, \ldots, x_n\}$, conventional PCA operates on zero-centered data obtained by diagonalizing the covariance matrix **Cov**:

$$\mathbf{Cov} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \tag{1}$$

In other words, PCA gives an eigen-decomposition of the covariance matrix:

$$\lambda \mathbf{V} = \mathbf{CovV} \tag{2}$$

3.2. Methods

Suppose the number of observations is n, the number of variables is p, and therefore, the size of input matrix $\mathbf{X}$ is $n \times p$. There are three main methods to perform the PCA:

1) Singular value decomposition (SVD), which is the canonical method for PCA [24].
2) Eigenvalue decomposition (EIG) of the covariance matrix [25]. The EIG algorithm is faster than SVD when the number of observations exceeds the number of variables, but it is less accurate because the condition number of the covariance is the square of the condition number of $\mathbf{X}$.
3) Alternating least squares (ALS) algorithm. It finds the best rank-k approximation by factoring input matrix into an $n \times k$ left factor matrix $\mathbf{L}$, and a $p \times k$ right factor matrix $\mathbf{R}$, where k is the number of principal components. The factorization uses an iterative method starting with random initial values. ALS is designed to better handle missing values. It can work well for data sets with a small percentage of missing data at random [26].

3.3. Illustration of OASIS Data

The GM images of each individual are of dimension $176 \times 208 \times 176$. We resampled the image to low-resolution of $22 \times 26 \times 22$. The undersampling [12, 27] is a conventional method. The low-resolution

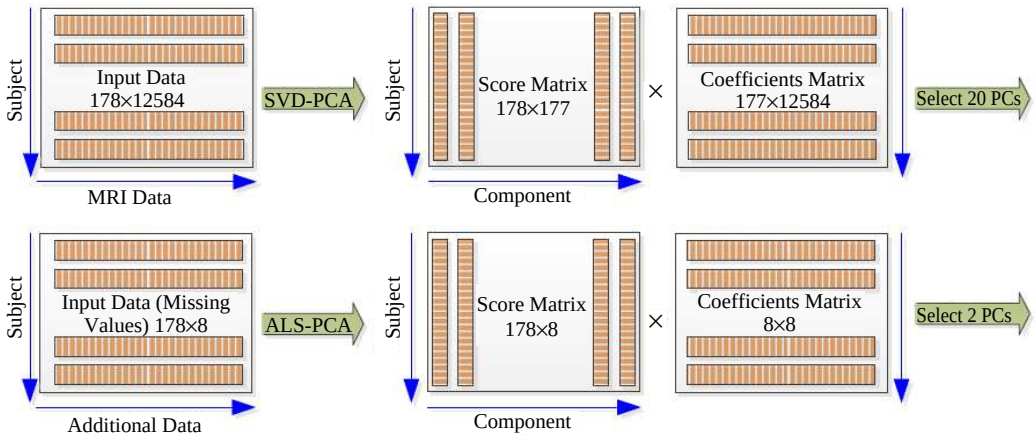

Figure 2. The SVD-PCA and ALS-PCA model for OASIS Data.

image was reformed to a row vector of 1×12584. The row vectors of 178 subjects were arranged into an 'input data matrix' with dimension 178×12584. Then, the input data matrix was decomposed into the principal component 'score matrix' and the 'coefficients matrix'. Here, the rows and columns of 'score matrix' correspond to subjects and components, respectively. Each row of the 'coefficient matrix' contains the coefficients for one principal component, and they are in the order of descending component variance.

The additional data of each subject contains 8 features (gender, age, education, SES, MMSE, eTIV, ASF, and nWBV). As a preprocessing, we used ALS algorithm for the feature reduction since there are missing values in the additional data. Figure 2 shows the PCA Model for OASIS Data.

4. METHODOLOGY OF KSVM-DT

The Support Vector Machine Decision Tree (SVM-DT) takes advantage of both the efficient computation of the tree architecture and the high classification accuracy of SVMs [28]. A multi-class problem can be divided into a series of two-class problem, which can be solved by SVMs, therefore SVMs can be used as nodes in SVM-DT to solve multi-class problem.

4.1. Skewness Tree

Two typical architectures of decision tree are skewness tree and normal tree [29]. Figure 3(a) shows a skewness tree of a four-class problem. The classifier SVM1 separates class 4 from other three classes. Next, SVM2 separates class 3 from class 1 & 2. Finally, the SVM3 separates class 1 from class 2. Figure 3(b) shows the corresponding normal tree of the same problem. SVM1 separates class 1 & 2 from class 3 & 4. The SVM2 separates class 1 from class 2, and the SVM3 separates class 3 from class 4. In this study, we choose the skewness tree as the skewness tree is superior to other structures [28, 30].

4.2. Kernel SVM

Traditional linear SVMs cannot separate intricately distributed practical data. In order to generalize it to nonlinear hyperplane, the kernel trick is applied to SVMs [31]. The kernel SVMs (kSVM) allows us to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space is a higher dimensional space. Though the classifier is a hyperplane

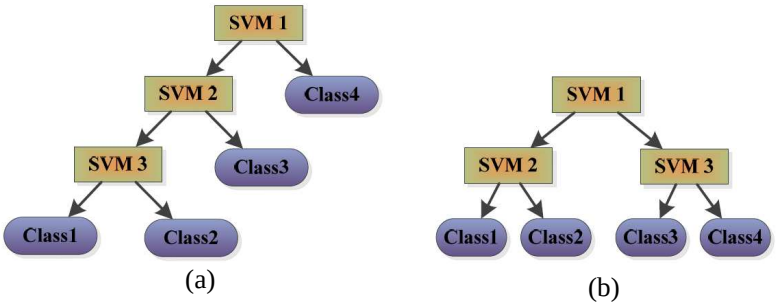

Figure 3. Illustrations of SVM-DT: (a) skewness tree and (b) normal tree of a four-class problem.

Table 4. Four common kernels.

Name	Formula	Parameter(s)
Homogeneous Polynomial	$k(x_i, x_j) = (x_i x_j)^d$	d
Inhomogeneous Polynomial	$k(x_i, x_j) = (x_i x_j + 1)^d$	d
Gaussian	$k(x_i, x_j) = \exp\left(-\frac{\|\|x_i - x_j\|\|^2}{2\sigma^2}\right)$	σ
Hyperbolic Tangent	$k(x_i, x_j) = \tanh(\kappa x_i x_j + c)$	κ, c

in the higher-dimensional feature space, it may be nonlinear in the original input space. Four common kernels [32] are listed in Table 4. For each kernel, there should be at least one adjusting parameter so as to make the kernel flexible and tailor itself to practical data. In this paper, we choose the Gaussian kernel due to its excellent performance [33].

4.3. PSO

Error penalty C is an important parameter in SVM [34]. If it is too large, we would have a high penalty for non-separable points, have to store too many support vectors, and may overfit. If it is too small, we may encounter underfitting. Traditional method uses trial-and-error to determine the optimal values of error penalty C and kernel parameter σ of kSVMs. It will cause heavy computation burden, and cannot guarantee to find the optimal or even near-optimal solutions. S. W. Fei, et al. [35] and C. L. Zhao, et al. [36] independently proposed to use PSO to optimize the parameters. The PSO is a populated global optimization method, derived from the research of the movement of bird flocking or fish schooling. It is easy and fast to implement. Besides, we introduced the cross validation to construct the fitness function used for PSO.

PSO performs searching via a swarm of particles which updates from iteration to iteration. To seek for the optimal solution, each particle moves in the direction of its previously best position (*pbest*) and the best global position in the swarm (*gbest*).

$$pbest_i = p_i(k^*) \ \ s.t.\ fitness\left(p_i(k^*)\right) = \min_{k=1,\ldots,t}\left[fitness\left(p_i(k)\right)\right] \tag{3}$$

$$gbest = p_{i^*}(k^*) \ \ s.t.\ fitness\left(p_{i^*}(k^*)\right) = \min_{\substack{i=1,\ldots,P \\ k=1,\ldots,t}}\left[fitness\left(p_i(k)\right)\right] \tag{4}$$

where i denotes the particle index, P denotes the total number of particles, k denotes the iteration index, and t denotes the current iteration number, and p denotes the position. The velocity and position of particles can be updated by the following equations.

$$v_i(t+1) = wv_i(t) + c_1 r_1\left(pbest_i(t) - p_i(t)\right) + c_2 r_2\left(gbest(t) - p_i(t)\right) \tag{5}$$

$$p_i(t+1) = p_i(t) + v_i(t+1) \tag{6}$$

where v denotes the velocity. The inertia weight w is used to balance the global exploration and local exploitation. The r_1 and r_2 are uniformly distributed random variables within range [0, 1]. The c_1 and c_2 are positive constant parameters called "acceleration coefficients".

It should be noted that only parameters C and kernel parameter σ are determined by PSO. The weights ω and biases b are still obtained by canonical quadratic programming (QP) method.

4.4. *K*-Fold Cross Validation

If the training set is used as validation set, then we will get an optimistically biased assessment. This estimate is called the in-sample estimate. Cross validation is a model validation technique for assessing how accurately the classifier will perform when generalizing to an independent set in practice, when an explicit validation set is not available [33]. The cross-validation estimate is called out-of-sample estimate.

We choose 5-fold cross-validation considering the best compromise between computational cost and reliable estimates. The dataset is randomly divided into 5 mutually exclusively subsets of approximately equal size, in which 4 subsets are used as training set S_T and the last subset is used as validation set S_V. The abovementioned procedure repeated 5 times, so each subset is used once for validation.

Suppose $i \in [1, 2, 3, 4, 5]$ denotes the run index of cross validation process. The whole dataset S is divided into five folds $[S_1, S_2, S_3, S_4, S_5]$. At ith run, the ith fold Si is set as the validation set $V(i)$, and the rest folds are set as the training set $T(i)$.

Figure 4 shows how PSO trains the kSVM-DT based on cross validation data. The weights/biases of kSVM are set as the variables, and the median square error (MSE) of the samples are set as the fitness function of PSO.

$$\min_{\omega,b} \sum\nolimits_{k \in Ti} \left\| K_k - P_k(\omega, b) \right\|^2 \tag{7}$$

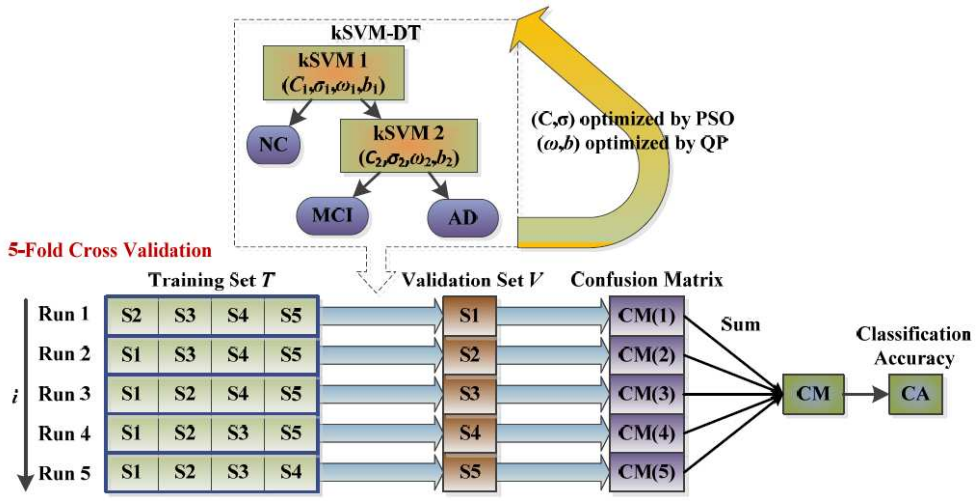

Figure 4. 5-fold cross validation data submitted to kSVM-DT optimized by PSO.

Here ω and b are weights/biases of kSVM, respectively. K and P denote the known class and predicted class of kth subject, respectively. PSO runs iteratively till the MSE of the training set is minimal. Afterwards, the validation set is submitted to the trained classifier so as to obtain the confusion matrix $CM(i)$ on validation set $V(i)$.

$$CM(i) = CM\left(K_k, P_k | k \in V(i)\right) \tag{8}$$

Recall that the known class K_k is obtained directly from $V(i)$, and the predicted class P_k is obtained from the output of the classifier that is trained on the training set $T(i)$. Since $V(i)$ and $T(i)$ are independent at each run, the confusion matrix $CM(i)$ of ith run reflects the out-of-sample error. In total, the $CM(i)$ of five runs are summed up to obtain the final confusion matrix. Considering

$$\bigcup_{i=1}^{5} V(i) = \bigcup_{i=1}^{5} S(i) = S \tag{9}$$

Therefore, the final confusion matrix checks all samples once and only once.

$$CM = \sum_{i=1}^{5} CM(i) = CM\left(K_k, P_k | k \in \bigcup\nolimits_{i=1}^{5} V(i)\right) = CM\left(K_k, P_k | k \in S\right) \tag{10}$$

The total classification accuracy CA is obtained by

$$CA = \sum\nolimits_{k \in S} \|K_k - P_k\|^2 \tag{11}$$

4.5. The Whole Process

The flowchart of our AD classification system is depicted in Figure 5. The detailed pseudocodes are listed as follows:

Step 1 Import.
- a. Import the OASIS dataset.
- b. Divide it into three types: MRI data, additional data, and target data.

Step 2 Feature Extraction and Reduction.
- a. Perform SVD-PCA on MRI data. Select 20 PCs according to variance explained.
- b. Perform ALS-PCA on Additional data. Select 2 PCs according to variance explained.
- c. Keep CDR data unchanged.
- d. Establish the extracted feature dataset as a 178×23 matrix.

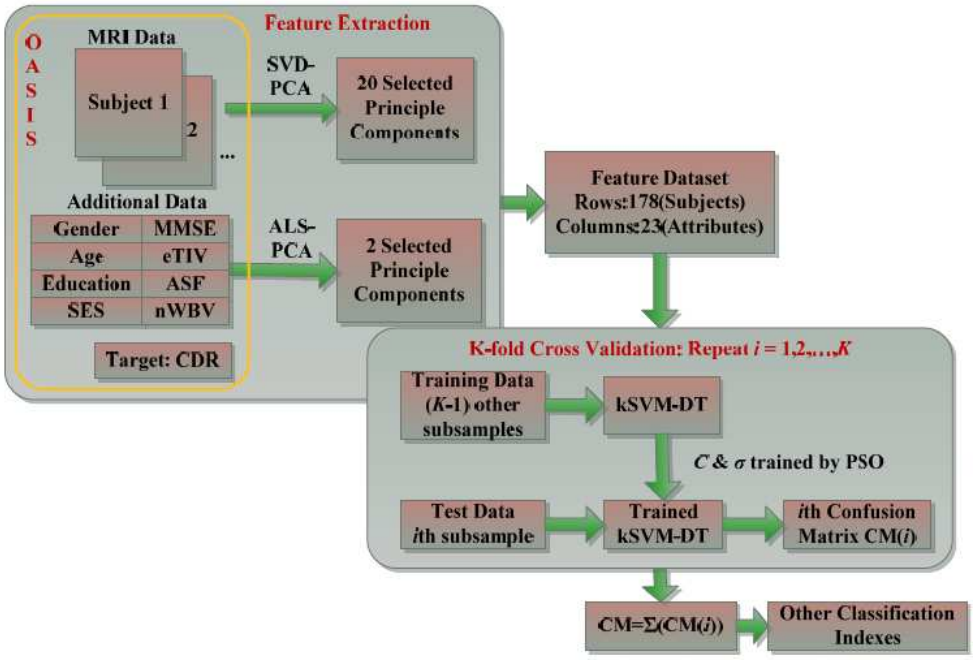

Figure 5. Flowchart of our AD classification system.

Step 3 K-fold Cross Validation.

a. Divide the 178 samples to five folds by stratified cross validation.
b. Let $i = 1$, and begin the ith run.
c. Select ith fold as the test data, and the rest four folds as the training data.
d. Training.
   i. Submit the training data to our kSVM-DT classifier model.
   ii. The parameters C & σ are trained by PSO.
   iii. The parameters ω and b are trained by QP algorithm.
e. Test.
   i. Submit the test data to the trained classifier.
   ii. Record the confusion matrix CM(i) of ith run.
f. If $i >= 5$ Jump to Step 4, otherwise return to Step 3c.

Step 4 Evaluation.

a. Sum up the confusion matrix of five runs to obtain the final CM.
b. Calculate the out-of-sample classification accuracy.

5. EXPERIMENTS AND DISCUSSIONS

The programs were in-house developed using Matlab 2013a, and run on IBM desktop with 3 GHz Intel i3 processor and 2 GB RAM.

5.1. PCA Results

The PCA results on MRI data and additional data are shown in Figure 6. The x-axis denotes the number of principal components, and the y-axis denotes the percentage of the total variance explained by each principal component. Figure 6(a) shows the result of MRI data. The 20 principal components accumulate to 14.1004% of the total variances. As we had resampled the original three dimensional (3D) image of $176 \times 208 \times 176$ to low-resolution of $22 \times 26 \times 22$, which eliminates the spatial dependence of 3D MRI data, the features after PCA are nearly independent to each other and the variance explained by

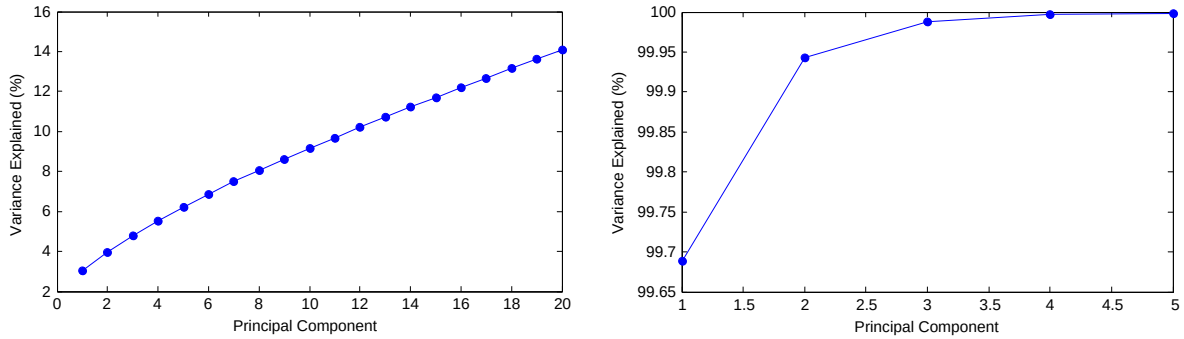

Figure 6. PCA of (a) MRI data and (b) additional data.

PCs are relatively low. Figure 6(b) shows the result of additional data, of which 2 principal components accumulate to 99.9453% of the total variances.

5.2. Cross Validation

We use the stratified cross validation method to divide the whole dataset into 5 folds. The detailed settings of five folds are shown in Table 5.

Table 5. Setting of five folds.

Fold	NC	MCI	AD	Total
1	19	12	4	35
2	19	12	5	36
3	20	11	5	36
4	20	11	5	36
5	19	11	5	35
Total	97	57	24	178

For ith run, we set ith fold as test set and other four folds as the training set. The confusion matrix of each run is shown in Table 6. The ith row and jth column in the confusion matrix denotes the count of observations known in group i but predicted to group j. The three classes in order are NC, MCI, and AD. The results of each run by SVM-DT and kSVM-DT are shown in Table 6.

Table 6. Confusion matrix of each run.

	Run 1	Run 2	Run 3	Run 4	Run 5	Sum
Training Set	$[S_2, S_3, S_4, S_5]$	$[S_1, S_3, S_4, S_5]$	$[S_1, S_2, S_4, S_5]$	$[S_1, S_2, S_3, S_5]$	$[S_1, S_2, S_3, S_4]$	
Validation Set	S_1	S_2	S_3	S_4	S_5	
Confusion Matrix (SVM-DT)	$\begin{bmatrix} 15 & 3 & 1 \\ 3 & 8 & 1 \\ 1 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 15 & 3 & 1 \\ 2 & 9 & 1 \\ 1 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 16 & 3 & 1 \\ 2 & 9 & 0 \\ 1 & 2 & 2 \end{bmatrix}$	$\begin{bmatrix} 16 & 3 & 1 \\ 3 & 7 & 1 \\ 1 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 15 & 3 & 1 \\ 2 & 9 & 0 \\ 1 & 2 & 2 \end{bmatrix}$	$\begin{bmatrix} 77 & 15 & 5 \\ 12 & 42 & 3 \\ 5 & 7 & 12 \end{bmatrix}$
Confusion Matrix (kSVM-DT)	$\begin{bmatrix} 16 & 3 & 0 \\ 2 & 9 & 1 \\ 0 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 16 & 2 & 1 \\ 2 & 10 & 0 \\ 1 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 17 & 3 & 0 \\ 2 & 8 & 1 \\ 0 & 1 & 4 \end{bmatrix}$	$\begin{bmatrix} 17 & 2 & 1 \\ 3 & 8 & 0 \\ 0 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} 17 & 2 & 0 \\ 2 & 9 & 0 \\ 1 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 83 & 12 & 2 \\ 11 & 44 & 2 \\ 2 & 6 & 16 \end{bmatrix}$

The total out-of-sample confusion matrix is obtained through summing the five run's results as shown in the last column of Table 6. We can see that for the proposed kSVM-DT method, 83 NC are classified correctly, but 12 NC are misclassified as MCI, and 2 NC are misclassified as AD. 44 MCI are classified correctly, but 11 MCI are misclassified as NC, and 2 MCI are misclassified as AD. 16 AD are classified correctly, but 2 AD are misclassified as NC, and 6 AD are misclassified as MCI. The overall classification accuracy of kSVM-DT is 80%

Contraversely, for the SVM-DT method, 77 NC are classified correctly, but 15 NC are misclassified as MCI, and 5 NC are misclassified as AD. 42 MCI are classified correctly, but 12 MCI are misclassified as NC, and 3 MCI are misclassified as AD. 12 AD are classified correctly, but 5 AD are misclassified as NC, and 7 AD are misclassified as MCI. The overall classification accuracy of SVM-DT is 74%.

The comparison results between SVM-DT and kSVM-DT are shown in Table 7. Here we can see the results of kSVM-DT are higher than those of SVM-DT in general. Using our proposed kSVM-DT method, the overall accuracy is 80%. For NC vs MCI, the sensitivity, specificity, and accuracy are 87%, 80%, and 85%, respectively. For NC vs AD, the sensitivity, specificity, and accuracy are 98%, 89%, and 96%. For MCI vs AD, the sensitivity, specificity, and accuracy are 96%, 73%, and 88%.

Table 7. Comparison between SVM-DT and kSVM-DT (SEN = Sensitivity, SPE = Specificity, ACR = Accuracy).

SVM-DT	NC vs MCI	NC vs AD	MCI vs AD	Overall
SEN	84%	94%	93%	
SPE	78%	71%	63%	
ACR	82%	90%	84%	74%
kSVM-DT	NC vs MCI	NC vs AD	MCI vs AD	
SEN	87%	98%	96%	
SPE	80%	89%	73%	
ACR	85%	96%	88%	80%

5.3. Performance of PSO

Take one fold run as example, the final parameters obtained by PSO were $C_1 = 172.5$, $\sigma_1 = 1.105$, $C_2 = 175.8$, $\sigma_2 = 1.098$. We compared PSO method with random selection method, which randomly

Table 8. Comparison between PSO and random selection method on kSVM-DT model.

Random Selection	C_1	σ_1	C_2	σ_2	Success Prediction	Overall ACR
Run 1	130.1	1.145	103.3	0.752	133	74.72%
Run 2	170.1	0.876	156.1	0.790	137	76.97%
Run 3	166.6	0.691	188.2	1.117	127	71.35%
Run 4	153.9	0.928	166.9	0.765	135	75.84%
Run 5	169.8	0.982	119.0	1.324	142	79.78%
Run 6	166.7	0.621	136.9	1.483	126	70.79%
Run 7	117.8	1.090	146.1	1.230	132	74.16%
Run 8	112.8	0.726	198.2	0.844	114	64.04%
Run 9	199.9	0.885	115.6	1.084	131	73.60%
Run 10	117.1	1.083	185.6	0.608	131	73.60%
Optimized by PSO	172.5	1.105	175.8	1.098	143	80%

generated the values of C in the range of [100, 200] and σ in the range of [0.5 1.5]. The results are shown in Table 8.

The classification accuracy varied with the change of parameters (C_1, σ_1, C_2, σ_2), so it was of importance to determine the optimal parameter values before training kSVM-DT. Data in Table 8 indicates that PSO is superior to random selection method.

5.4. Computation Time

We ran our program 10 times, and the averaged computation time of every stage is listed in Table 9. The preprocessing cost the most time as 45 minutes, since the total dataset is 48 GB. Afterwards, the PCA on MRI data and additional data cost 1.084 s and 0.008 s, respectively. Finally, the train and classification of kSVM-DT expends 2.356 s and 0.022 s, respectively. In practice, we always construct the classifier in advance, and instruct the examiners the skills of using the classifier. Therefore, it costs about 0.022 s to get the computer-aided diagnosis for each new patient.

Table 9. Averaged computation time.

Stage	Time
Preprocessing	45 min
PCA on MRI data	1.084 s
PCA on additional data	0.008 s
kSVM-DT Train	2.356 s
kSVM-DT Classification	0.022 s

6. CONCLUSIONS

In this paper, we proposed a hybrid classification system for distinguishing NC, MCI, and AD based on structural MRI images. We used MRI data, demographics, clinical examination, and derived anatomic volume as the training data. The CDR was used as the target data. We used PCA to reduce the dimensionality of the feature vectors of the MRI data and the resultant principal components retained important information. The kSVM-DT method gathered the principal components from MRI data and additional data, and the final classification accuracy is 80%. Considering that the problem is three-class classification, the result is relatively outstanding. Our method can be used as an auxiliary tool for diagnosis.

The first limitation of our method is that the classifier establishes machine-oriented rules not human-oriented rules. Technicians cannot understand what the weights/biases of the kSVM-DT mean. Therefore, it gives us a research direction to build human-oriented model. Another limitation arises as how to increase the classification accuracy. Current studies indicate AD is associated with metabolites in the gray matter; however, what we use in this research is structural MRI, which may not cover the sufficient information containing the cause of AD. Adding other imaging techniques, such as magnetic resonance spectroscopy imaging (MRSI) measuring the metabolites in the brain, may increase the classification accuracy of the classifier.

The contributions of the paper are: 1) combining MRI data with additional data to improve classification accuracy; 2) using ALS PCA algorithm to process missing data; 3) the use of the kSVM-DT method; 4) determining the optimal parameter C and σ by PSO; 5) using cross validation to obtain the out-of-sample error estimate. The future tentative work will focus on the following aspects: 1) to extract more efficient features; 2) try other classifiers such as artificial neural network, Bayesian classifier, and hidden Markov models; 3) to include the phase image of structural MRI data; 4) to add the MRSI data.

ACKNOWLEDGMENT

The work is supported by the Nanjing Normal University Research Foundation for Talented Scholars (No. 2013119XGQ0061) and the National Natural Science Foundation of China (No. 40871176, No. 610011024). Besides, the authors express their gratitude of the OASIS dataset that comes from NIH grants P50AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382 and R01 MH56584.

REFERENCES

1. Hahn, K., et al., "Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease — Revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence," *NeuroImage*, Vol. 81, 96–109, 2013.
2. Brookmeyer, R., et al., "Forecasting the global burden of Alzheimer's disease," *Alzheimers Dement*, Vol. 3, No. 3, 186-191, 2007.
3. Chen, X., W. Yang, and X. Huang, "ICA-based classification of MCI vs HC," *2011 Seventh International Conference on Natural Computation (ICNC)*, Vol. 3, 1658–1662, 2011.
4. Kubota, T., Y. Ushijima, and T. Nishimura, "A region-of-interest (ROI) template for three-dimensional stereotactic surface projection (3D-SSP) images: Initial application to analysis of Alzheimer disease and mild cognitive impairment," *International Congress Series*, Vol. 1290, 128–134, 2006.
5. Pennanen, C., et al., "Hippocampus and entorhinal cortex in mild cognitive impairment and early AD," *Neurobiology of Aging*, Vol. 25, No. 3, 303–310, 2004.
6. Lee, W., B. Park, and K. Han, "Classification of diffusion tensor images for the early detection of Alzheimer's disease," *Computers in Biology and Medicine*, Vol. 43, No. 10, 1313–1320, 2013.
7. López, M. M., et al., "SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA," *Neuroscience Letters*, Vol. 464, No. 3, 233–238, 2009.
8. Camacho, J., J. Picó, and A. Ferrer, "Corrigendum to 'the best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?'," *Anal. Chim. Acta*, Vol. 642, 59–68, 2009; *Analytica Chimica Acta*, Vol. 658, No. 1, 106–106, 2010.
9. Ortiz, A., et al., "LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease," *Pattern Recognition Letters*, Vol. 34, No. 14, 1725–1733, 2013.
10. Ardekani, B. A., K. Figarsky, and J. J. Sidtis, "Sexual dimorphism in the human corpus callosum: An MRI study using the OASIS brain database," *Cereb Cortex*, Vol. 10, No. 25, 2514–2520, 2012.
11. Ardekani, B. A., et al., "Corpus callosum shape changes in early Alzheimer's disease: An MRI study using the OASIS brain database," *Brain Struct. Funct.*, Vol. 219, No. 1, 343–352, 2013.
12. Bin Tufail, A., et al. "Multiclass classification of initial stages of Alzheimer's disease using structural MRI phase images," *2012 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 317–321, 2012.
13. "What is OASIS? OASIS: Cross-sectional MRI data in young, middle aged, nondemented and demented older adults 2013," Available from: http://www.oasis-brains.org/.
14. Möller, C., et al., "Different patterns of gray matter atrophy in early- and late-onset Alzheimer's disease," *Neurobiology of Aging*, Vol. 34, No. 8, 2014–2022, 2013.
15. Alexander, G. E., et al., "Gray matter network associated with risk for Alzheimer's disease in young to middle-aged adults," *Neurobiology of Aging*, Vol. 33, No. 12, 2723–2732, 2012.
16. Smith, S. M., "Fast robust automated brain extraction," *Human Brain Mapping*, Vol. 17, No. 3, 143–155, 2002.
17. Kuslansky, G., et al., "Detecting dementia with the Hopkins verbal learning test and the mini-mental state examination," *Archives of Clinical Neuropsychology*, Vol. 19, No. 1, 89–104, 2004.
18. Maxeiner, H. and M. Behnke, "Intracranial volume, brain volume, reserve volume and

morphological signs of increased intracranial pressure — A post-mortem analysis," *Legal Medicine*, Vol. 10, No. 6, 293–300, 2008.

19. Buckner, R. L., et al., "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume," *NeuroImage*, Vol. 23, No. 2, 724–738, 2004.
20. Fotenos, A. F., et al., "Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD," *Neurology*, Vol. 64, No. 6, 1032–1039, 2005.
21. Williams, M. M., et al., "Progression of Alzheimer's disease as measured by clinical dementia rating sum of boxes scores," *Alzheimer's & Dementia*, Vol. 9, No. 1, S39–S44, 2013.
22. Marcus, D. S., et al., "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cogn. Neurosci.*, Vol. 19, No. 9, 1498–1507, 2007.
23. Zhang, Y. and L. Wu, "An MR brain images classifier via principal component analysis and kernel support vector machine," *Progress In Electromagnetics Research*, Vol. 130, 369–388, 2012.
24. Gass, S. I. and T. Rapcsák, "Singular value decomposition in AHP," *European Journal of Operational Research*, Vol. 154, No. 3, 573–584, 2004.
25. Rajendra Acharya, U., et al., "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Systems with Applications*, Vol. 39, No. 10, 9072–9078, 2012.
26. Kuroda, M., et al., "Acceleration of the alternating least squares algorithm for principal components analysis," *Computational Statistics & Data Analysis*, Vol. 55, No. 1, 143–153, 2011.
27. Cuingnet, R., et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, Vol. 56, No. 2, 766–781, 2011.
28. Arun Kumar, M. and M. Gopal, "A hybrid SVM based decision tree," *Pattern Recognition*, Vol. 43, No. 12, 3977–3987, 2010.
29. Xu, Z., P. Li, and Y. Wang, "Text classifier based on an improved SVM decision tree," *Physics Procedia*, Vol. 33, 1986–1991, 2012.
30. Nasseri, M., H. Tavakol-Davani, and B. Zahraie, "Performance assessment of different data mining methods in statistical downscaling of daily precipitation," *Journal of Hydrology*, Vol. 492, 1–14, 2013.
31. Acevedo-Rodríguez, J., et al., "Computational load reduction in decision functions using support vector machines," *Signal Processing*, Vol. 89, No. 10, 2066–2071, 2009.
32. Deris, A. M., A. M. Zain, and R. Sallehuddin, "Overview of support vector machine in modeling machining performances," *Procedia Engineering*, Vol. 24, 308–312, 2011.
33. Zhang, Y. and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors*, Vol. 12, No. 9, 12489–12505, 2012.
34. Wu, K.-P. and S.-D. Wang, "Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space," *Pattern Recognition*, Vol. 42, No. 5, 710–717, 2009.
35. Fei, S.-W., "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Systems with Applications*, Vol. 37, No. 10, 6748–6752, 2010.
36. Zhao, C., et al., "Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine," *Expert Systems with Applications*, Vol. 38, No. 8, 9908–9912, 2011.