

Research Article

Discrimination between Alzheimer's Disease and Mild Cognitive Impairment Using SOM and PSO-SVM

Shih-Ting Yang,¹ Jiann-Der Lee,¹ Tzyh-Chyang Chang,^{1,2} Chung-Hsien Huang,¹
Jiun-Jie Wang,³ Wen-Chuin Hsu,^{4,5} Hsiao-Lung Chan,¹ Yau-Yau Wai,^{3,6} and Kuan-Yi Li⁷

¹ Department of Electrical Engineering, Chang Gung University, Tao-Yuan 333, Taiwan

² Department of Occupational Therapy, Bali Psychiatric Center, New Taipei City 249, Taiwan

³ Department of Medical Imaging and Radiological Sciences, Chang Gung University, Tao-Yuan 333, Taiwan

⁴ Department of Neuroscience, Chang Gung Memorial Hospital, Tao-Yuan 333, Taiwan

⁵ Chang Gung Dementia Center, Chang Gung Memorial Hospital, Tao-Yuan 333, Taiwan

⁶ Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital, Tao-Yuan 333, Taiwan

⁷ Department of Occupational Therapy, Chang Gung University, Tao-Yuan 333, Taiwan

Correspondence should be addressed to Jiann-Der Lee; jdlee@mail.cgu.edu.tw

Received 15 February 2013; Accepted 13 April 2013

Academic Editor: Chung-Ming Chen

Copyright © 2013 Shih-Ting Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, an MRI-based classification framework was proposed to distinguish the patients with AD and MCI from normal participants by using multiple features and different classifiers. First, we extracted features (volume and shape) from MRI data by using a series of image processing steps. Subsequently, we applied principal component analysis (PCA) to convert a set of features of possibly correlated variables into a smaller set of values of linearly uncorrelated variables, decreasing the dimensions of feature space. Finally, we developed a novel data mining framework in combination with support vector machine (SVM) and particle swarm optimization (PSO) for the AD/MCI classification. In order to compare the hybrid method with traditional classifier, two kinds of classifiers, that is, SVM and a self-organizing map (SOM), were trained for patient classification. With the proposed framework, the classification accuracy is improved up to 82.35% and 77.78% in patients with AD and MCI. The result achieved up to 94.12% and 88.89% in AD and MCI by combining the volumetric features and shape features and using PCA. The present results suggest that novel multivariate methods of pattern matching reach a clinically relevant accuracy for the a priori prediction of the progression from MCI to AD.

1. Introduction

Alzheimer's disease (AD) [1] is the most common type of dementia. Clinical signs are characterized by progressive cognitive deterioration, together with declining activities of daily living and by neuropsychiatric symptoms or behavioral changes. The early detection of AD is potentially challenging because of several reasons. First of all, there existed no known biomarkers. The disease usually has an insidious onset which can be a combination of genetic and environmental factors. It is difficult to differentiate other types of dementia.

Mild cognitive impairment (MCI) is a transitional stage between normal aging and demented status. The syndrome

is defined by the greater cognitive decline than age and education matched individuals, but no interference of daily function [2]. According to the major symptoms, MCI is characterized with memory loss and cognitive impairment. Research has reported that MCI has a risk between 10% to 64% developing AD [3, 4]. AD is a progressively neurodegenerative disorder and is distinguished from MCI by the progressive deterioration of daily function. The prevalence of AD increases dramatically at age 65 and it affects approximately 26 million people worldwide, which may increase fourfolds by the year of 2050. Recent reports in the treatment or prevention of AD lead to a growing concerns in the early diagnosis. Therefore, the detection of changes in

brain tissues that reflect the pathological processes of MCI would prevent or postpone the disease progresses either from normal control to MCI or from MCI to AD. If MCI can be diagnosed at an early stage and effectively intervened, then it is possible to reduce the advanced damages.

Since the poor performance in memory and execution function indicates the high risk of dementia, the probable AD patients are usually evaluated by standardized neuropsychological tests [5–8]. Additionally, many studies have been proposed to examine the predictive abilities of nuclear imaging with respect to AD and other dementia illnesses [9–13]. However, under the consideration of imaging cost and noninvasive requirement, magnetic resonance imaging (MRI) has been widely used for early detection and diagnosis of MCI and AD [14–17].

Atrophy typically starts in the medial temporal and limbic areas, subsequently extending to parietal association areas, and finally to frontal and primary cortices. Early changes in hippocampus and entorhinal cortex have been demonstrated with the help of MRI, and these changes are consistent with the underlying pathology of MCI and AD. Many studies have used manual or automatic methods to measure hippocampus and entorhinal cortex [18–20]. Hippocampal volumes and entorhinal cortex measures have been found to be equally accurate in distinguishing between AD and normal cognitive elderly subjects [21]. However, the segmentation and identification of hippocampus or entorhinal cortex are usually sensitive to the subjective opinion of the operator and also time consuming. In addition, the enlargement of ventricles is also a significant characteristic of AD due to neuronal loss. Ventricles are filled with cerebrospinal fluid (CSF) and surrounded by gray matter (GM) and white matter (WM). As a result, by measuring the ventricular enlargement, hemispheric atrophy rate shows higher correlation with the disease progression.

In this study, we have designed an MRI-based classification framework to distinguish the patients of MCI and AD from normal individuals using multiple features and different classifiers. Since the features adopted here are volume-related and shape-related, we also aimed to investigate whether the combination of both statistical analysis and principal component analysis (PCA) would improve the accuracies of classification than using volume-related alone, shape-related alone, or all features. Our hypothesis was that the combination of all MRI-based features is helpful for distinguishing the patients with early Alzheimer’s disease from the subjects with mild cognitive impairment and healthy controls, respectively.

The remainder of this paper is organized as follows. Section 2 illustrated the proposed scheme, including features extraction and used classifiers, that is, self-organizing map (SOM), support vector machine (SVM), particle swarm optimization (PSO), and the proposed hybrid PSO-SVM. Statistical analysis, experimental results, and discussion are revealed in Section 3. Finally, conclusions are included in Section 4.

2. The Proposed Schemes

Figure 1 is the flowchart that demonstrated the system we proposed. In the step of *Feature Extraction*, spatial normalization

is performed by coregistering the brain MRI data from each individual to a T1-weighted MRI template such that these images of the investigated subjects will be in the same scale space. Next, with the aids of segmentation and morphological procedures, all MRI brain images are segmented into GM, WM, CSF, and ventricle’s tissues and shape descriptors. Here, volume-related and shape-related features are utilized for further classification. The step of *Feature Reduction* is divided into two parts: (1) Mann-Whitney U test is adopted to filter out the features with low discriminative power; (2) principal component analysis (PCA) is applied to reduce the dimensions of feature space. Route I only uses U test; Route II is combined with U test and PCA. At last, a classifier, for example, SOM, SVM, and PSO-SVM, is employed to classify tested volunteers into three categories: normal individuals, MCI, and AD patients. The details of the proposed method are described below.

2.1. Spatial Normalization of MRI Data. Spatial normalization of the brain images is useful for determining what happens generically over individuals. It is a procedure to register an MRI data set to a standard coordinate system, also known as Talairach and Tournoux coordinate system [22]. With the aid of normalization, all images were spatially normalized to stereotactic space ICBM-152 [23] via a 12-degrees-of-freedom affine transformation which normalizes the brain in terms of dimensions, position, and spatial orientation.

2.2. Volume Features Extraction. The volumes of brain tissues such as GM, WM, and CSF indicate important information, especially in brain degeneration diseases [24]. A clustering-based segmentation algorithm provided by SPM8 [25] is using a modified Gaussian mixture model to extract GM, WM and CSF probability maps from whole-brain MRI data. The intensities of voxels belonging to each of these clusters conform to a normal distribution which can be described by a mean, a variance, and the number of voxels belonging to the distribution. Here, the volumes of GM, WM, CSF, and whole-brain are calculated by

$$\begin{aligned} \text{volume}_{\text{tissue}} &\approx \sum_{\forall i \in I} (P(C_{\text{tissue}} | f(i)) > 0.5), \\ \text{volume}_{\text{whole}} &\approx \sum_{\forall i \in I} (P(C_{\text{GMVWM}} | f(i)) > 0.5), \end{aligned} \quad (1)$$

where i is any pixel of the MRI data and $f(i)$ stands for the gray level of i . C means the cluster. tissue stands for the parts of GM, WM, or CSF. Figure 2 illustrates the segmentation results of the normal individual and AD patient used in this study.

Next, we employ region growing and double threshold algorithm [26] to extract binary ventricle volume data, that is, $M(x, y, z)$. The morphological operators, for example, erosion and dilation, are used to obtain the binary ventricle regions. And the edges of binary images are detected by applying Sobel operation on a slice-by-slice basis. Then, this segmented region will construct a binary mask image. In

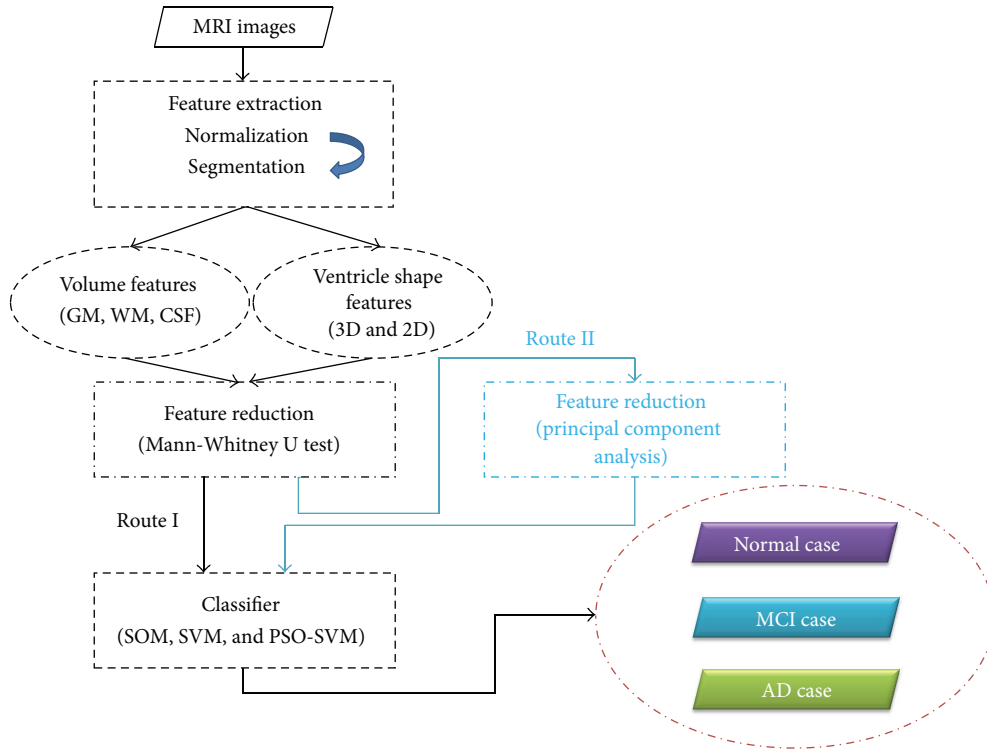


FIGURE 1: Flowchart of the proposed image-aided diagnosis system.

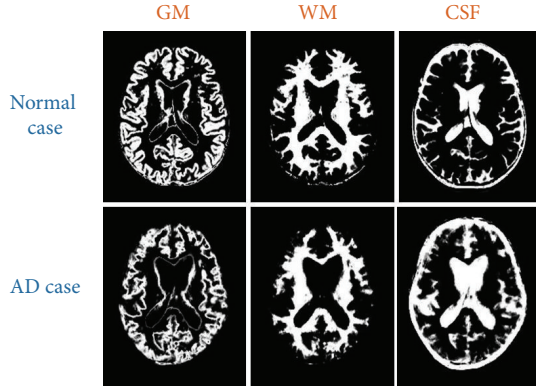


FIGURE 2: Segmentation results of a normal individual and an AD patient used in this study.

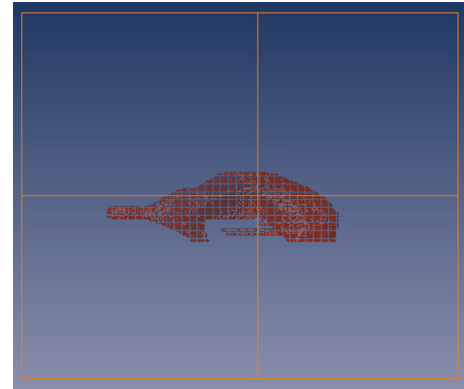


FIGURE 3: Sagittal view of segmented ventricle.

this mask image, 1 (white) denotes the ventricle pixel, and 0 (black) denotes the nonventricle pixel. Finally, we can calculate the volume of cerebral ventricle by

$$\text{volume}_{\text{ventricle}} \approx \sum_{i \in M} (P(C_{\text{ventricle}} | f(i)) = 1), \quad (2)$$

where i is any pixel of the mask data, M is the mask image, and $f(i)$ denotes the gray level of i .

2.3. Shape Features Extraction. The volume features, which are extracted from the whole three dimensional volume, cannot capture the variation of the anatomical shape. Wang

et al. [27, 28] proposed a ventricle shape-based method for improved classification of Alzheimer's patients. Therefore, to enhance the accuracy of the classification, in addition to the volume features, we also added ventricle shape features. Figure 3 shows the sagittal view of ventricle that we segmented. The shape features we analyzed are composed of two types: three-dimensional shape features and two-dimensional shape features. The algorithms to obtain these features are illustrated in the following subsections.

2.3.1. 3D Shape Features. To obtain the feature of 3D shape, a leave-one-out method is used to construct training set and

testing set following Wang's method. Three sets of probability map were then built using

$$P_t(x, y, z) = \frac{1}{N} \sum_{i=1}^N I_t^i(x, y, z), \quad (3)$$

where t indicates the type of the subjects, inclusive of normal control, AD, and MCI. N is the number of training samples, and I denotes the gray level of the ventricular mask image. In order to compare the differences of patients (AD and MCI) and normal controls, we subtracted the normal probability map from the patient probability map to obtain the discriminate map. At last, a matching coefficient (MC) between a testing input and the discriminate map is calculated by

$$\begin{aligned} \text{MC}_{\text{Normal or patient}}^i &= \sum_{\forall x, y, z} D(x, y, z) \\ &\times T_{\text{Normal or patient}}^i(x, y, z), \end{aligned} \quad (4)$$

where $D(x, y, z)$ is the discriminate map and T denotes the testing ventricular mask image.

2.3.2. 2D Shape Features. The 2D shape features are extracted from the segmented ventricles on a slice-by-slice basis. In 2D viewpoint, there are many 2D ventricle slices for each case. In order to effectively compare the differences in each case, we selected the slices with maximum areas from 3D ventricle data sets as the datum plane. These 2D shape features used herein are referred to the work of Yang et al. [29] and listed as follows: (1) *Area*, (2) *Perimeter*, (3) *Compactness*, (4) *Elongation*, (5) *Rectangularity*, (6) *Distances*, (7) *Minimum thickness*, and (8) *Mean signature value*.

2.4. Learning Methods for Classification. Machine learning algorithms can be organized into a taxonomy based on the desired outcome of the algorithm or the type of input available during training the machine. They are often divided into supervised, unsupervised, and reinforcement learning (RL). Supervised learning requires the explicit provision of input-output (I/O) pairs and the task is one of constructing a mapping from one to the other. Non-supervised learning has no concept of target data and performs processing only on the input data. In contrast, RL uses a scalar reward signal to evaluate I/O pairs and hence discover, through trial and error, the optimal outputs for each input. In this sense, RL can be thought of as intermediary to supervised and non-supervised learning since some form of supervision is present, albeit in the weaker guise of the reward signal. As such, the trained algorithm may be treated as a "black box" encapsulating knowledge gleaned from the training data whose inputs are useful for producing the expected outcome. For this reason, machine learning and computer-aided diagnostics (CADs) have been of growing interest in the field of medical applications. To evaluate whether the performance of supervised and non-supervised methods is good or not, we used three classifiers to produce the outcome.

In many researches of pattern recognition, dataset is often divided into two subsets of training and testing. The former

is used to create the model, and the latter is used to assess the accuracy of the model to predict the unknown sample. This method can be called Train-and-Test method. Cross-validation is the experimental method to effectively estimate the generalization error. In this study, leave-one-out cross-validation (LOOCV) is adopted in three classifiers to estimate dependable generalization error. LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. In this section, the classifiers we adopted are illustrated in the following subsections particularly.

2.4.1. Self-Organizing Map Architecture. A self-organizing map (SOM) is a type of artificial neural network for the visualization of high-dimensional data. In general, SOMs are divided into two parts: training and mapping. Training builds the map using input examples, called a Kohonen map [30]. An SOM consists of components called nodes or neurons. Each node has a set of neighbors. When this node wins a competition, not only its weight is adjusted, but those of the neighbors are also changed. They are not changed as much though. The further the neighbor is from the winner, the smaller its weight change. Furthermore, as training goes on, the neighborhood gradually shrinks. At the end of training, the neighborhoods have shrunk to zero size.

When a training example is fed to the network, its Euclidean distance to all weight vectors is computed by using (5). Here n denotes the dimension of data, and t is the index of the data item in a given sequence,

$$x(t) = \{\zeta_1(t), \zeta_2(t), \dots, \zeta_n(t)\}. \quad (5)$$

The neuron with weight vector most similar to the input is called the best matching unit (BMU). The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector. The magnitude of the change decreases with time and with distance from the BMU. The update formula for a neuron with weight vector is

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t) [x(t) - m_i(t)], \quad (6)$$

where $\alpha(t)$ is a monotonically decreasing learning coefficient and $x(t)$ is the input vector. The neighborhood function $h_{ci}(t)$ depends on the lattice distance between the BMU and neuron. The neighborhood function $h_{ci}(t)$ is

$$h_{ci}(t) = \frac{e^{-\|r_i - r_c\|^2}}{2\sigma^2(t)}. \quad (7)$$

Figure 4 illustrates the procedure of SOM classifier. In this study, we use a two-stage method for learning [31]. First, we adopt less iterative time, higher learning rate, and large neighborhood distance for learning and make it convergence speedily. After repeating many times, we can acquire network parameters which have the best convergence. Next, combining higher iterative time, less learning rate, and small neighborhood distance with network parameters obtained in first stage to conduct second learning and adjust network parameters slowly. At last, we obtain these parameters:

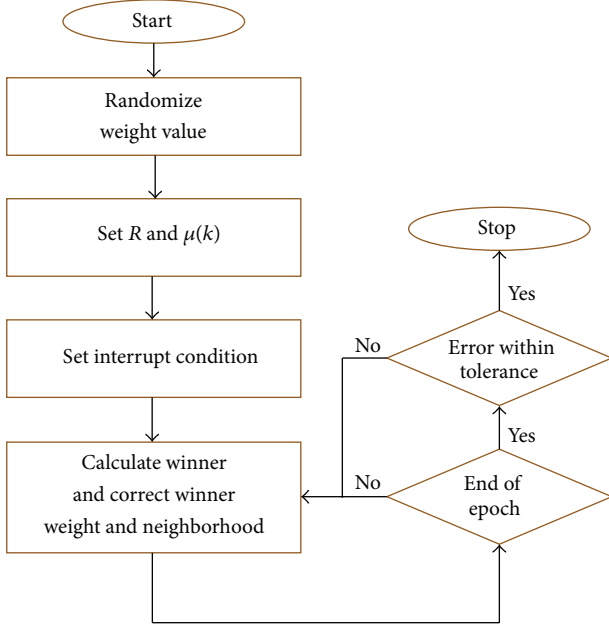


FIGURE 4: Basic procedure of SOM classifier.

iterative time is set as 1000 epochs, ordering phase learning rate = 0.9, tuning phase learning rate = 0.5, and tuning phase neighborhood distance = 0.5. In order to verify the stability of SOM to generalize the correct tendency, the classifier was trained 10 times to get reliable results. Thirty cases are chosen (AD = 7, Normal = 7, MCI = 8) to be the training set randomly. Scaling of variables is of special importance in our model since the SOM algorithm uses Euclidean metric to measure distances between vectors. In order to solve this problem, we achieved this by linearly scaling all variables so that their variances were equal to one.

2.4.2. Support Vector Machine. SVM is a type of artificial neural networks that is, trained by using supervised learning, have shown their advantage on reducing training-and-testing errors, resulting in obtaining higher recognition accuracy [32]. However, some feature data are linearly nonseparable. In some situations, features are not perfectly separable, especially at the border between categories. To allow some flexibility in separating the categories, SVMs utilize a cost parameter, denoted as C , to control the trade-off between allowing training errors and forcing rigid margins. The cost function with C is defined as (8), where ζ_i is a slack variable,

$$\text{Cost} = C \sum_{i=1}^N (\zeta_i). \quad (8)$$

Mapping the patterns in a high dimension feature space is generated through combining features to form a kernel matrix. The kernel matrix is usually constructed by using a kernel function which takes two patterns as arguments and outputs a value. In this study, a radial basis function (RBF) kernel, as shown in (9), is employed. We use one-against-rest assemblies classifiers that distinguish one from all

TABLE I: Demographic data and cognitive scores.

Group	Normal control	MCI	AD
Individuals (male/female)	17 (10/7)	18 (9/9)	17 (9/8)
Mean age (yrs)	71.43 ± 4.43	72.50 ± 4.00	72.70 ± 3.93
Education time (yrs)	10.17 ± 5.21	8.22 ± 5.25	5.24 ± 5.36
MMSE scores	28.18 ± 1.70	25.06 ± 4.11	13.29 ± 6.69

the other classes. This strategy consists of constructing one SVM per class, which is trained to distinguish the samples of one class from the samples of all remaining classes. Usually, classification of an unknown pattern is done according to the maximum output among all SVMs,

$$k(x_i, y_j) = e^{-\gamma \|x_i - y_j\|^{\text{Fit}_p}}, \quad i = j = 1, 2, \dots, n, \quad (9)$$

where x_i denotes the input vector, y_j denotes the j th prototype vector, and $\text{Fit}_p = \text{correctly-classified}/\text{total number of testing data}$. Finally, the optimal solution can be solved by using Lagrange method,

$$L_p \equiv \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i - \sum_{i=1}^m \alpha_i \{y_i (w \cdot x_i + b) - 1 + \zeta_i\},$$

$$L_D \equiv \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j y_i y_j^k (x_i, y_j), \quad (10)$$

where $\|w\|$ is the Euclidean norm of w , α_i that stands for the Lagrange multipliers, L_p is the Lagrange function, and L_D is the dual solution of L_p . C and γ are used to control the trade-off between training errors and generalization ability in SVM with RBF kernel. Therefore, a PSO was utilized to find the optimal combination of C and γ .

2.4.3. Hybrid PSO-SVM. Particle swarm optimization (PSO) algorithm [33, 34] uses particles moving in an m -dimensional space to search solutions of an optimization problem with m variables. In our approach, PSO is initialized and searches for the optimal particle iteratively. Each particle represents a candidate solution. SVM classifier is built for each candidate solution to evaluate its performance. Velocity and position of particles can be updated by

$$v_{ij}^{t+1} = w \cdot v_{ij}^t + c_1 \text{rand}_1(\text{pbest}_{ij}^t - x_{ij}^t) + c_2 \text{rand}_2(\text{gbest}_{ij}^t - x_{ij}^t)$$

$$x_{ij}^{t+1} = x_{ij}^t + v_{ij}^{t+1}, \quad (11)$$

where t is evolutionary generation, v_{ij} is the velocity of particle i on dimension j , and x_{ij} stands for the position of particle i on dimension j . Inertia weight w is used to balance the global exploration and local exploitation, rand_1 and rand_2 are random functions, and c_1 and c_2 are personal and social learning factors. As we know, if the number of particles,

TABLE 2: Statistical analysis of features.

Features	Mean volume \pm SD				
	Normal	MCI	AD	<i>P</i> value (NC versus MCI)	<i>P</i> value (NC versus AD)
Volume					
V_{GM}	862.4 \pm 42.7	824.6 \pm 57.8	789.7 \pm 84.3	0.016	0.007
V_{WM}	637.6 \pm 45.8	601.8 \pm 21.2	558.1 \pm 63.4	0.021	0.019
V_{CSF}	863.1 \pm 112.9	909.7 \pm 128.5	971.8 \pm 132.5	0.038	0.017
Shape					
Area	1792.4 \pm 278.5	1903.5 \pm 426.6	2361.1 \pm 802.3	0.029	0.024
Area (PR)	673.5 \pm 121.5	874.9 \pm 132.5	911.4 \pm 183.2	0.024	0.018
Area (PL)	647.1 \pm 137.2	872.5 \pm 142.5	910.9 \pm 183.5	0.031	0.011
Area (FR)	151.9 \pm 117.6	231.5 \pm 162.4	262.4 \pm 167.8	0.020	0.009
Area (FL)	162.7 \pm 91.0	258.2 \pm 144.3	278.5 \pm 189.2	0.022	0.010
Perimeter	226.7 \pm 23.1	276.9 \pm 20.2	289.8 \pm 27.6	0.029	0.019
Circularity	45.6 \pm 4.9	39.8 \pm 3.6	38.2 \pm 2.7	0.039	0.021
Elongation	1.1 \pm 0.4	1.4 \pm 0.6	1.3 \pm 0.2	0.016	0.009
Rectangularity	0.5 \pm 0.2	0.6 \pm 0.4	0.6 \pm 0.1	0.028	0.016
$d(A, G)$	37.3 \pm 2.1	38.4 \pm 3.7	40.6 \pm 4.2	0.031	0.037
$d(B, G)$	36.1 \pm 1.8	39.2 \pm 3.1	43.1 \pm 6.1	0.034	0.028
$d(C, G)$	38.6 \pm 4.3	41.4 \pm 2.9	42.9 \pm 4.6	0.042	0.030
$d(D, G)$	34.7 \pm 2.9	39.7 \pm 1.4	42.8 \pm 4.1	0.022	0.028
$d(A, C)$	72.8 \pm 4.3	81.7 \pm 8.4	83.8 \pm 8.4	0.009	0.011
$d(B, D)$	72.5 \pm 4.9	78.2 \pm 3.1	81.6 \pm 8.2	0.011	0.007
Min thickness	27.4 \pm 3.8	29.0 \pm 2.6	30.1 \pm 3.4	0.020	0.009
Mean Sig.	25.6 \pm 3.1	27.9 \pm 2.7	29.8 \pm 3.1	0.032	0.013

TABLE 3: PCs and their proportion of total variation.

Features	No. of principal component								
	Proportion (%)	C1	C2	C3	C4	C5	C6	C7	C8
Volume features (3)		64.16*	31.57*	4.27					
Shape features (17)		48.79*	23.39*	9.43*	6.45*	3.28*	2.13*	1.01*	0.73*
Volume + shape (20)		49.31*	19.98*	13.62*	6.93*	4.47*	2.35*	0.99	0.72

denoted as P , is too large, it might cause the optimization process to be time consuming. On the contrary, if P is too small, then it is hard to find the optimal solution due to the limited search area. In the literature [35], it is proven that the optimal solution can be obtained when P is between 20 and 40. In this work, the number of the iterations and P is set to 200 and 30, respectively. Similarly, the parameters c_1 , c_2 , and w will affect the convergence of optimization process. If they are set too large, it causes the particle velocity to be speedy and thus cannot obtain the optimal solution. On the other hand, it is time consuming to find the optimal solution [36]. Therefore, we set c_1 , c_2 , and w to 2, 2, and 0.8, respectively.

More specifically, based on the approach [37], the proposed hybrid PSO-SVM aims at optimizing the accuracy of SVM classifier by randomly generating the parameters (C and γ) and estimating the best values for regularization of kernel parameters for SVM model. Basic operation of hybrid PSO-SVM proposed in this paper is given in Figure 5.

This process continues until the performance of SVM converges. The termination criteria are that the iteration

number reaches the maximum number of iterations (100%) or the value of global optimal fitness does not improve after 200 consecutive iterations. In this study, 22 cases were chosen (AD = 7, Normal = 7, MCI = 8) to be the training set.

3. Experimental Results and Discussion

3.1. Materials. According to the research [4], most patients with Alzheimer's disease are aged 65 years or older. Therefore, most of the subjects in the whole data we choose are over 65 years old. The image data used in this study were provided by Chang Gung Memorial Hospital, Lin-Kou, Taiwan. The degree of clinical severity for each participant was evaluated by experienced clinicians whom conducted independent semistructured interviews which included a set of questions regarding the functional status of the participant, along with a standardized neurologic, psychiatric, and health examinations. This interview generates an overall Clinical Dementia Rating (CDR) and Mini Mental State Examination

TABLE 4: Classification results (SOM).

Proportion	Volume features	Volume features + PCA	Shape features	Shape features + PCA	Volume + shape features	Volume + shape features + PCA
AD (versus NC)						
Accuracy	76.47%	82.35%	64.71%	70.59%	76.47%	88.24%
Sensitivity	81.25%	87.50%	68.75%	70.59%	76.47%	88.24%
Specificity	77.78%	83.33%	66.67%	70.59%	76.47%	88.24%
MCI (versus NC)						
Accuracy	61.11%	66.67%	50.00%	50.00%	66.67%	72.22%
Sensitivity	78.57%	85.71%	64.29%	64.29%	75.00%	86.67%
Specificity	66.67%	71.43%	57.14%	57.14%	68.42%	75.00%

TABLE 5: Confused matrix with SOM (volume + shape/volume + shape + PCA).

	NC	MCI	AD
NC	13/15	2/2	1/0
MCI	3/2	12/13	3/5
AD	1/0	4/3	13/15

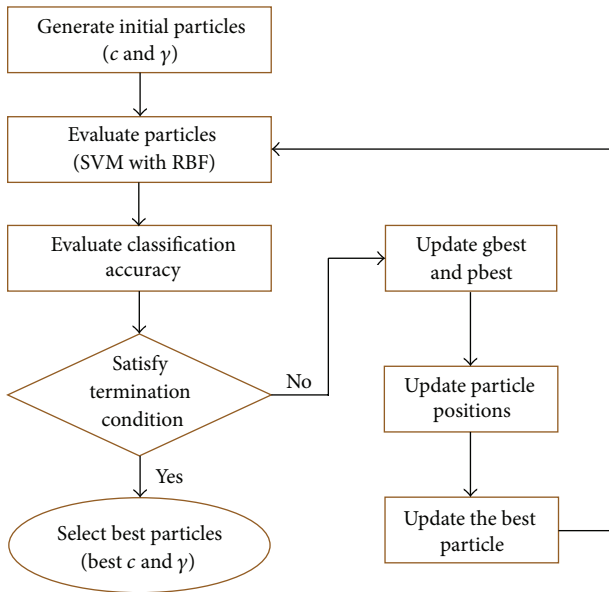


FIGURE 5: Basic operation of proposed PSO-SVM approach.

(MMSE) score. The whole dataset consists of three groups comprising normal control, MCI, and AD. Demographic information is provided in Table 1.

The whole-brain MRI scans were obtained by a 3T MR scanner (Trio A TIM system, Siemens, Erlangen, Germany). T1-weighted images were acquired by magnetization-prepared 180 degrees radio-frequency pulses and rapid gradient-echo (T1-MPRAGE) series. The following imaging parameters were used: repetition time (TR) = 2000 ms, echo time (TE) = 4.16 ms, and flip angle = 9 degrees. The results

were represented as a 224×256 matrix, and slice thickness = 1 mm in 160 slices.

3.2. Statistical Analysis and Classification. Through image processing techniques, we obtained individual volume and shape features. In order to confirm whether there is a significant effect of the classification for these features, we use statistical MW test to compare differences between three groups on various features (continuous variables).

The MW test, also called a Mann-Whitney U or Mann-Whitney Wilcoxon test, is a nonparametric rank-based test for identifying the difference between populations with respect to their medians or means. The test does not require sample data to be normal (sample > 30), and it is relatively insensitive to the nonhomogeneity of the variance of sample data. The null hypothesis is that the two populations from which samples have been drawn have equal medians or means. The alternatives are that the populations do not have equal medians. The two samples are combined, and all sample observations are ranked from smallest to largest. It was performed on each feature to evaluate its discriminative power, as shown in (12). U_{obt} is the smaller value taken from the sum of U_1 and U_2 , where n_1 and n_2 are the sizes of the first and second samples, respectively,

$$Z_U = \frac{U_{\text{obt}} - (n_1 n_2 / 2)}{\sqrt{n_1 n_2 (n_1 + n_2 + 1) / 12}}. \quad (12)$$

The P values obtained from the tests can provide the probability that a variation would assume a value greater than or equal to the observed value strictly by chance. It is known that the P value which is less than the predetermined significance level (0.05) would result in the rejection of the null hypothesis at the 5% (significance) level. All statistical results of volume and shape features we adopted (< 0.05) are shown in Table 2, inclusive of three volume features and seventeen shape features.

3.3. Results. Although the features we adopted have statistical significance (< 0.05) between three groups, some of the features may be redundant or have high correlation. Therefore, principal component analysis (PCA) [38] is used to reduce the dimensionality of a data set consisting of a

TABLE 6: Classification results (SVM).

Proportion	Volume features	Volume features + PCA	Shape features	Shape features + PCA	Volume + shape features	Volume + shape features + PCA
AD (versus NC)						
Accuracy	70.59%	70.59%	58.82%	64.71%	76.47%	82.35%
Sensitivity	70.59%	66.67%	66.67%	78.57%	76.47%	87.50%
Specificity	70.59%	68.75%	63.16%	70.00%	76.47%	83.33%
MCI (versus NC)						
Accuracy	55.56%	61.11%	44.44%	50.00%	77.78%	83.33%
Sensitivity	66.67%	64.71%	61.54%	75.00%	77.78%	88.24%
Specificity	60.00%	61.11%	54.55%	60.87%	76.47%	83.33%

TABLE 7: Confused matrix with SVM (volume + shape/volume + shape + PCA).

	NC	MCI	AD
NC	13/15	0/1	1/0
MCI	3/2	14/15	3/3
AD	1/0	4/2	13/14

large number of interrelated variables, while retaining as much as possible of the variation present in the data set. On the other hand, it can also improve the computation time required for classification. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated and are ordered so that the first few retain most of the variation present in all of the original variables. In order to effectively represent all the data, we used the PCs that captured 95% total variation in data set. To train a volume-feature-based classification, the first two principal components were adopted. To train a shape-feature-based classification, only the first eight principal components were adopted. When we integrated volume and shape features into classification, the first six principal components were used to stand for all of the features. Table 3 gives the variances and the coefficients of the PCs, when the analysis is done on the correlation matrix. The symbol * indicates that this PCA coefficient is used as a feature for classification. SOM, SVM, and PSO-SVM were used to train a classifier, and the results were presented in Tables 4, 5, 6, 7, 8, and 9.

It showed the results of accuracy (proportion of all subjects correctly classified), sensitivity (proportion of individuals with a true positive result), and specificity (proportion of individuals with a true negative result) when using different features. The derivations of accuracy, sensitivity, and specificity were expressed in (13), where TP = true positive, TN = true negative, and FP = false positive. Obviously, incorporating shape features, volume features, and PCA provided excellent classification ability than using only one of them,

$$\text{Accuracy (ACC)} = \frac{(\text{TP} + \text{TN})}{(P + N)}$$

Sensitivity or true positive rate (TPR)

$$= \frac{\text{TP}}{P} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

Specificity or True Negative Rate (TNR)

$$= \frac{\text{TN}}{N} = \frac{\text{TN}}{(\text{FP} + \text{TN})}$$

(13)

3.4. Discussion. In this study, we investigated the feasibility of using anatomical MR images to extract different types of features as a predictive marker for AD and MCI. We employed multiple features and different classifiers to identify the patients with AD and MCI from normal participants. From the results, volumetric analysis, inclusive of gray/white matter, cerebrospinal fluid, and local shape analysis on ventricle, provides significant atrophy information. Especially, the properties of gray matter volume, ventricular area, elongation, mean signature value, and distances show the statistical significance (<0.01). This implies that using the volume and shape features have the potential ability to identify normal control, AD, and MCI.

By combining both the volumetric features and shape features, the classification accuracy of SOM reached up to 76.47% and 66.67% in patients with AD and MCI, respectively. Moreover, with the help of PCA algorithm, the classification result was improved up to 88.24% and 72.22% in patients with AD and MCI, respectively. The classification accuracy of SVM reached up to 76.47% and 77.78% in patients with AD and MCI, respectively. Moreover, with the help of PCA algorithm, the classification result was improved up to 82.35% and 83.33% in patients with AD and MCI, respectively. With the hybrid classification framework based on PSO, the result achieved up to 82.35% and 77.78% in AD and MCI. Moreover, with the help of PCA algorithm, the classification result was improved up to 94.12% and 88.89% in patients with AD and MCI, respectively. According to the results, combining PSO-SVM with statistical analysis and principal component analysis (PCA) would improve the accuracy of classification.

It was also noted that the classification ability was significant for AD and normal control than the patients with

TABLE 8: Classification results (PSO-SVM).

Proportion	Volume features	Volume features + PCA	Shape features	Shape features + PCA	Volume + shape features	Volume + shape features + PCA
AD (versus NC)						
Accuracy	76.47%	76.47%	70.59%	76.47%	82.35%	94.12%
Sensitivity	76.47%	76.47%	70.59%	76.47%	87.50%	94.12%
Specificity	76.47%	76.47%	70.59%	76.47%	83.33%	94.12%
MCI (versus NC)						
Accuracy	66.67%	66.67%	55.56%	50.00%	77.78%	88.89%
Sensitivity	75.00%	75.00%	66.67%	69.23%	87.50%	94.12%
Specificity	68.42%	68.42%	60.00%	59.09%	78.95%	88.88%

TABLE 9: Confused matrix with PSO-SVM (volume + shape/volume + shape + PCA).

	NC	MCI	AD
NC	15/16	1/1	0/0
MCI	2/1	14/16	3/1
AD	0/0	3/1	14/16

MCI. MCI is a transitional stage between normal cognitive aging and dementia. Therefore, the characteristics of patients with MCI were similar to AD subjects. On the other hand, the characteristic of patients with MCI was also possibly similar to normal participants. Combination with other features was essential to improve the accuracy of classification ability for patients with MCI in an early stage.

4. Conclusion

In this paper, we compared different methods for the classification of patients with AD and MCI based on anatomical T1-weighted MRI. To evaluate and compare the performances of each method, two classification experiments were performed: CN versus AD and CN versus MCI. It is observed that the volume features and shape features can be integrated to increase classification accuracy with the low computational complexity. Classification results also verify our hypothesis that the combination of multimodal features, including volume and shape features, outperforms a single modality of features, possibly because different features are mutually complementary. Furthermore, it is proven that statistical analysis and PCA can achieve accuracies significantly better than all the features that are adopted. In the performance of classifiers used here, it is shown that PSO-SVM can achieve the best accuracy, sensitivity, and specificity, no matter for CN versus AD and CN versus MCI.

For the moment, the classified results are greater for patients with AD and normal participants than for patients with MCI. It can provide clinically useful information at the large-scale population-based screening studies. The results would be welcomed for prognosticating disease progression and providing an objective evaluation of cognitive rehabilitation treatments for dementing illness.

Acknowledgments

The work was supported by National Science Council, Taiwan, under Grant no. NSC98-2221-E-182-040-MY3 and Chang Gung Memorial Hospital with Grant no. CMRP D270053.

References

- [1] S. Gauthier, B. Reisberg, M. Zaudig et al., "Mild cognitive impairment," *The Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.
- [2] D. M. Geslani, M. C. Tierney, N. Herrmann, and J. P. Szalai, "Mild cognitive impairment: an operational definition and its conversion rate to Alzheimer's disease," *Dementia and Geriatric Cognitive Disorders*, vol. 19, no. 5-6, pp. 383–389, 2005.
- [3] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *Journal of Internal Medicine*, vol. 256, no. 3, pp. 183–194, 2004.
- [4] *Alzheimer's disease facts and figures*, 2012, http://www.alz.org/alzheimers_disease_facts_figures.asp?type=homepage.
- [5] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini mental state: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.
- [6] C. P. Hughes, L. Berg, and W. L. Danziger, "A new clinical scale for the staging of dementia," *British Journal of Psychiatry*, vol. 140, no. 6, pp. 566–572, 1982.
- [7] E. L. Teng, K. Hasegawa, A. Homma et al., "The cognitive abilities screening instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia," *International Psychogeriatrics*, vol. 6, no. 1, pp. 45–58, 1994.
- [8] T. N. Tombaugh, "Trail Making Test A and B: normative data stratified by age and education," *Archives of Clinical Neuropsychology*, vol. 19, no. 2, pp. 203–214, 2004.
- [9] P. Padilla, J. M. Górriz, J. Ramírez et al., "Analysis of SPECT brain images for the diagnosis of Alzheimer's disease based on NMF for feature extraction," *Neuroscience Letters*, vol. 479, no. 3, pp. 192–196, 2010.
- [10] R. Chaves, J. Ramirez, J. M. Gorriz, and C. G. Puntonet, "Alzheimer's Disease Neuroimaging Initiative, Association rule-based feature selection method for Alzheimer's disease diagnosis," *Expert Systems With Applications*, vol. 39, no. 14, pp. 11766–11774, 2012.
- [11] J. Ramírez, J. M. Górriz, F. Segovia et al., "Computer aided diagnosis system for the Alzheimer's disease based on partial

- least squares and random forest SPECT image classification,” *Neuroscience Letters*, vol. 472, no. 2, pp. 99–103, 2010.
- [12] A. Gallix, J. M. Gorriz, J. Ramirez, I. A. Illan, and E. W. Lang, “On the empirical mode decomposition applied to the analysis of brain SPECT images,” *Expert Systems With Applications*, vol. 39, no. 18, pp. 13451–13461, 2012.
- [13] D. Salas-Gonzalez, J. M. Gorriz, J. Ramirez et al., “Two approaches to selecting set of voxels for the diagnosis of Alzheimer’s disease using brain SPECT images,” *Digital Signal Processing*, vol. 21, pp. 746–755, 2012.
- [14] J. E. Iglesias, J. Jiang, C. Y. Liu, and Z. Tu, “Alzheimer’s Disease Neuroimaging Initiative, Classification of Alzheimer’s disease using a self-smoothing operator,” in *Proceedings of the 14th International Conference on Medical Image Computing and Computer Assisted Intervention*, 2012.
- [15] P. Vemuri, J. L. Gunter, M. L. Senjem et al., “Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies,” *NeuroImage*, vol. 39, no. 3, pp. 1186–1197, 2008.
- [16] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, “Multimodal classification of Alzheimer’s disease and mild cognitive impairment,” *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [17] P. Vemuri, H. J. Wiste, S. D. Weigand et al., “MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change,” *Neurology*, vol. 73, no. 4, pp. 294–301, 2009.
- [18] K. Juottonen, M. P. Laakso, K. Partanen, and H. Soininen, “Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease,” *American Journal of Neuroradiology*, vol. 20, no. 1, pp. 139–144, 1999.
- [19] O. Colliot, G. Chételat, M. Chupin et al., “Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus,” *Radiology*, vol. 248, no. 1, pp. 194–201, 2008.
- [20] J. H. Morra, Z. Tu, L. G. Apostolova et al., “Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer’s disease, mild cognitive impairment, and elderly controls,” *NeuroImage*, vol. 45, no. 1, pp. S3–S15, 2009.
- [21] K. Kantarci, “Magnetic resonance markers for early diagnosis and progression of Alzheimer’s disease,” *Expert Review of Neurotherapeutics*, vol. 5, no. 5, pp. 663–670, 2005.
- [22] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain*, Thieme, New York, NY, USA, 1988.
- [23] J. Mazziotta, A. Toga, A. Evans et al., “A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM),” *Philosophical Transactions of the Royal Society B*, vol. 356, no. 1412, pp. 1293–1322, 2001.
- [24] K. H. Fritzsche, A. von Wangenheim, D. D. Abdala, and H. P. Meinzer, “A computational method for the estimation of atrophic changes in Alzheimer’s disease and mild cognitive impairment,” *Computerized Medical Imaging and Graphics*, vol. 32, no. 4, pp. 294–303, 2008.
- [25] UCL Institute of Neurology, <http://www.fil.ion.ucl.ac.uk/spm/>.
- [26] C. F. Jiang, C. H. Huang, and S. T. Yang, “Using maximal cross-section detection for the registration of 3D image data of the head,” *Journal of Medical and Biological Engineering*, vol. 31, no. 3, pp. 217–226, 2011.
- [27] J. Wang, A. Ekin, and G. De Haan, “Shape analysis of brain ventricles for improved classification of alzheimer’s patients,” in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP ’08)*, pp. 2252–2255, October 2008.
- [28] J. Wang, G. De Haan, D. Unay, O. Soldea, and A. Ekin, “Voxel-based discriminant map classification on brain ventricles for Alzheimer’s disease,” in *Medical Imaging: Image Processing*, vol. 7259 of *Proceedings of SPIE*, February 2009.
- [29] S. T. Yang, J. D. Lee, C. H. Huang, J. J. Wang, W. C. Hsu, and Y. Y. Wai, “An image-aided diagnosis system for dementia classification based on multiple features and self-organizing map,” *Lecture Notes in Computer Science*, vol. 6444, no. 2, pp. 462–469, 2010.
- [30] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [31] S. Wu and T. W. S. Chow, “Self-organizing and self-evolving neurons: a new neural network for optimization,” *IEEE Transactions on Neural Networks*, vol. 18, no. 2, pp. 385–396, 2007.
- [32] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [33] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, December 1995.
- [34] Z. Cui, L. Wang, and Y. Tan, “Particle swarm optimization with active congregation,” *ICIC Express Letters*, vol. 4, no. 4, pp. 1167–1172, 2010.
- [35] M. Kudo and J. Sklansky, “Comparison of algorithms that select features for pattern classifiers,” *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [36] Y. Shi and R. Eberhart, “Modified particle swarm optimizer,” in *Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC ’98)*, pp. 69–73, May 1998.
- [37] C. J. Tu, L. Y. Chuang, J. Y. Chang, and C. H. Yang, “Feature selection using PSO-SVM,” *IAENG International Journal of Computer Science*, vol. 33, no. 1, pp. 138–143, 2007.
- [38] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2002.