J. Shanghai Jiaotong Univ. (Sci.), 2014, 19(5): 555-560
DOI: 10.1007/s12204-014-1540-4

# Discriminant Analysis in the Study of Alzheimer's Disease Using Feature Extractions and Support Vector Machines in Positron Emission Tomography with $^{18}$F-FDG

*SU Sai-sai*[1] (苏赛赛), *CHEN Ke-wei*[1,2] (陈克伟), *HUANG Qiu*[1*] (黄　秋)
(1. School of Biomedical Engineering, Shanghai Jiaotong University, Shanghai 200030, China;
2. Banner Alzheimer's Institute, Phoenix AZ85006, USA)



**Abstract:** With more successful applications of advanced medical imaging technologies in clinical diagnosis, various analytic discriminant approaches, by seeking the imaging based characteristics of a given disease to achieve automatic diagnosis, gain greater attention in the medical community. However the existing computer-aided discriminant procedures for Alzheimer's disease (AD) are yet to be improved for better identifying patients with mild cognitive impairment (MCI) from those with AD and those who are cognitively normal. In this work we present a computer assisted diagnosis approach by first statistically extracting characteristics from whole brain 2-deoxy-2-($^{18}$F)fluoro-D-glucose positron emission tomography ($^{18}$F-FDG PET) images, and then using support vector machines for classification. Evaluations of the proposed procedure with patient data exhibit satisfactory accuracies in distinguishing AD from its early stage MCI, and normal controls.
**Key words:** Alzheimer's disease, feature extraction, classification, mild cognitive impairment (MCI)
**CLC number:** R 319　　**Document code:** A

## 0 Introduction

Alzheimer's disease (AD) is the most common cause of dementia in elder people, affecting approximately 5.2 million individuals in the United States of America according to the Alzheimer's association latest report[1]. It is predicted that, the total estimated prevalence is about 13.8 million by 2050, with the growth rate about a million new cases per year[2]. Besides, those with mild cognitive impairment (MCI) have an increased risk of eventually developing Alzheimer's or another type of dementia, are very difficult to diagnose due to the lack of symptoms in this stage[3-4]. Though there is no effective treatment currently for AD or MCI, early adequate diagnosis is critically important for possible new drug development to slow down the progression and for identification of other causes which might be treatable.

In the last 20 years, with the development of medical imaging techniques, researchers developed a range of measurements for the scientific study as well as clinical evaluation of AD by using brain imaging[5-7]. The 2-deoxy-2-($^{18}$F)fluoro-D-glucose positron emission tomography ($^{18}$F-FDG PET) is currently one of the most characterized brain imaging techniques, information of which reflects the brain metabolism of glucose, the main energy for the proper function of human brains[8]. It has been confirmed that glucose metabolic reduction in the parietal-temporal, frontal and posterior cingulate cortices is related to AD[9].

Various AD diagnosis approaches can be found in the literatures[10-15]. Among these, most methods are based on the analysis of region of interest (ROI) followed by some discriminant functions. Others are using some statistical analysis tools such as statistical parametric mapping (SPM) and 3D stereotactic surface projections (3D-SSP)[16-17], to perform voxel-wise statistical analyses. Meanwhile, dimensionality reduction and feature selection based methods are gaining more attention by neuroimaging scientists. Instead of using the brain voxel-based data directly, these methods extract typically just handful features from images, and use these features as inputs to future statistical discriminant analysis for clinical diagnostic purposes.

In this study, we attempted to distinguish AD patients from those who are at the early stage of disease, MCI, and from normal controls by using ROI based, principal component analysis (PCA) based and linear discriminant analysis (LDA) based methods for feature extraction and support vector machine (SVM)

**Received date:** 2013-06-10
**Foundation item:** the National Natural Science Foundation of China (No. 81201114), the Shanghai Municipal Natural Science Foundation (No. 11ZR1416700), and the Innovation Program of Shanghai Municipal Education Commission (No. 13ZZ017)
∗**E-mail:** qiuhuang@sjtu.edu.cn

for classification. Different from traditional ROI based method, ROIs in our approach could be selected automatically. We also used an improved PCA method and an LDA method for feature extractions. The proposed method, tested on 375 Alzheimer's disease neuroimaging initiative (ADNI) positron emission tomography (PET) data, has satisfactory accuracy in distinguish AD from MCI and normal control (NC) when comparing with existing methods.

As AD research focuses more and more on early abnormality detection and intervention many years before the onset of the disease, identifying sensitive biomarkers becomes critical. As intermediate stage of normal aging and onset of the disease, proper MCI diagnosis is the focus of many studies and also very important to move AD research to preclinical (i.e., before the onset of the disease) stage. It is with this background, we attempted to evaluate a combination of statistical procedures with the hypothesized statistical power increase. In this regard, the study is more medical challenge motivated rather than purely for the development of new methodologies.

## 1 Materials and Preprocessing

### 1.1 Subjects

In this work, we used $^{18}$F-FDG PET data acquired for ADNI (http://adni.loni.usc.edu/) due to the small size of samples currently available in China with the standardized protocol. We downloaded data of 375 ADNI1 $^{18}$F-FDG PET subjects acquired using PET scanners by various manufactures (such as General Electric (GE), Philips and Siemens) and categorized them into three classes shown in Table 1, where $P_{\mathrm{m}}$ is the percentage of male. The mini-mental state examination (MMSE) scores and the clinical dementia rating (CDR) are also listed.

**Table 1 Information of the data collected**

| Subject | $P_{\mathrm{m}}$/% | Age | MMSE | CDR |
|---|---|---|---|---|
| AD ($n = 123$) | 61 | $77.3 \pm 7.4$ | $20.5 \pm 1.5$ | $0.7 \pm 0.2$ |
| MCI ($n = 133$) | 65 | $76.5 \pm 7.2$ | $24.0 \pm 1.7$ | $0.5 \pm 0.0$ |
| NC ($n = 119$) | 55 | $75.0 \pm 6.7$ | $29.0 \pm 1.1$ | $0.0 + 0.0$ |

We performed three binary classification tasks in our study after grouping these data as follows:

**Group 1** Consisting of data of AD subjects and NC subjects, for the task to distinguish AD from NC, with AD subjects labeling as positive and normal controls as negative.

**Group 2** Consisting of AD (positive) and MCI (negative) data, for the task to distinguish AD from MCI.

**Group 3** Consisting of MCI (positive) and NC (negative) data, for the task to distinguish MCI from NC.

### 1.2 Image Preprocessing

All images were normalized to Montreal Neurological Institute (MNI) template using SPM8 software (www.fil.ion.ucl.ac.uk/spm/software/spm8), through a general affine model with 12 parameters. After the linear normalization, the intensity normalization was performed to account for the global count variation by normalizing the intensity to the mean intensity value of the whole brain.

## 2 Method

### 2.1 Feature Extractions

After $^{18}$F-FDG PET images are preprocessed, dimensionality reduction and feature selection are performed to lower the complexity and increase the feasibility of the analysis. In this work we conducted three different approaches of feature extractions: ROI, LDA and PCA based. For LDA and PCA approaches, we adopted an adaptive threshold to choose dominant eigenvalues via the Fisher discriminant ratio (FDR).

#### 2.1.1 *ROI Based Method*

We selected on the basis of anatomical automatic labeling (AAL) brain regions of interest which are related to AD hypometabolism. The related brain regions, composed of 22 AAL ROIs in total, are Frontal_Mid_L, Frontal_Mid_R, Frontal_Sup_L, Frontal_Sup_R, Temporal_Mid_L, Temporal_Mid_R, Temporal_Inf_L, Temporal_Inf_R, Fusiform_L, Fusiform_R, Hippocampus_L, Hippocampus_R, Parietal_Inf_L, Parietal_Inf_R, Cingulum_Post_L, Cingulum_Post_R, Occipital_Sup_L, Occipital_Sup_R, Occipital_Mid_L, Occipital_Mid_R, Occipital_Inf_L, Occipital_Inf_R[18].

Meanwhile, instead of selecting ROIs related to AD according to literatures, we proposed another ROI selection approach which is based on statistical analysis. We calculated the mean intensity value of each of the 116 AAL ROIs, and performed $t$-test for each of the three tasks described above. To be compatible with the AAL based ROI approach, we also selected 22 ROIs using the between-group significance (the first 22 most significant ROIs).

#### 2.1.2 *LDA Based Method*

Alternatively, we chose LDA and PCA methods to select features correlated to AD. LDA is a classic technique for data classification and dimensionality reduction which reshapes the scatter in data in order to make it more reliable for classification. Based on the classic LDA algorithm, we used an improved method[19] designed for 2D matrices. The method not only solved the problem of small sample size, but also made the features typical for discrimination by performing two projections simultaneously in both vertical and horizontal directions on the matrix.

With the preprocessed images, we first reshaped all

$M_x \times M_y \times M_z$ image volumes to $M \times M$ matrices, where $M$ is the largest previous integer of the square root of $M_x \times M_y \times M_z$. Then we computed the between-class scatter matrix ($\boldsymbol{S}_{\mathrm{B}}$) and within-class scatter matrix ($\boldsymbol{S}_{\mathrm{W}}$) of the $N$ matrices for training in a specific group. In this work, a group contained two classes of data, for example Group 1 for AD and NC. And we used 70% data downloaded for training and 30% for validation. The scatter matrices are defined as follows:

$$\boldsymbol{S}_{\mathrm{B1}} = \sum_{i=1}^{C} N_i (\boldsymbol{\mu}_i^{\mathrm{LDA}} - \boldsymbol{\mu}^{\mathrm{LDA}})(\boldsymbol{\mu}_i^{\mathrm{LDA}} - \boldsymbol{\mu}^{\mathrm{LDA}})^{\mathrm{T}}, \quad (1)$$

$$\boldsymbol{S}_{\mathrm{W1}} = \sum_{i=1}^{C} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij}^{\mathrm{LDA}} - \boldsymbol{\mu}_i^{\mathrm{LDA}})(\boldsymbol{x}_{ij}^{\mathrm{LDA}} - \boldsymbol{\mu}_i^{\mathrm{LDA}})^{\mathrm{T}}, \quad (2)$$

$$\boldsymbol{S}_{\mathrm{B2}} = \sum_{i=1}^{C} N_i (\boldsymbol{\mu}_i^{\mathrm{LDA}} - \boldsymbol{\mu}^{\mathrm{LDA}})^{\mathrm{T}}(\boldsymbol{\mu}_i^{\mathrm{LDA}} - \boldsymbol{\mu}^{\mathrm{LDA}}), \quad (3)$$

$$\boldsymbol{S}_{\mathrm{W2}} = \sum_{i=1}^{C} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij}^{\mathrm{LDA}} - \boldsymbol{\mu}_i^{\mathrm{LDA}})^{\mathrm{T}}(\boldsymbol{x}_{ij}^{\mathrm{LDA}} - \boldsymbol{\mu}_i^{\mathrm{LDA}}), \quad (4)$$

where $\boldsymbol{S}_{\mathrm{B}}$ and $\boldsymbol{S}_{\mathrm{W}}$ represent the between-class and within-class scatter matrices, and the subscripts 1 and 2 of $\boldsymbol{S}_{\mathrm{B}}$ and $\boldsymbol{S}_{\mathrm{W}}$ are corresponding to the row-direction and column-direction of pixel-matrix. As mentioned above, $N_i$ is the number of training subjects in the $i$th class. Parameter $\boldsymbol{\mu}^{\mathrm{LDA}}$ represents the whole mean of the dataset while $\boldsymbol{\mu}_i^{\mathrm{LDA}}$ is the $i$th class mean value, $\boldsymbol{x}_{ij}^{\mathrm{LDA}}$ represents the $j$th subject in the $i$th class, and $C$ is the class number (in our case $C = 2$). Note both $\boldsymbol{x}_{ij}^{\mathrm{LDA}}$ and $\boldsymbol{\mu}^{\mathrm{LDA}}$ are matrices now. For each direction, we can construct a matrix $\boldsymbol{S}$ with the scatter matrixes $\boldsymbol{S}_{\mathrm{B}}$ and $\boldsymbol{S}_{\mathrm{W}}$ as:

$$\boldsymbol{S} = \boldsymbol{S}_{\mathrm{W}}^{-1} \boldsymbol{S}_{\mathrm{B}}. \quad (5)$$

Eigenvectors of $\boldsymbol{S}$ were calculated and then an adaptive threshold (see Subsection 2.4 below) was adopted to choose $m$ leading eigenvectors according to corresponding eigenvalues, where $m < M$ could be different for different groups. With one set of eigenvectors for each direction, we constructed two projection matrices $\boldsymbol{Z}$ and $\boldsymbol{X}$. Finally we projected the image to:

$$\boldsymbol{P}_k = \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{x}_k^{\mathrm{LDA}} \boldsymbol{X}, \quad (6)$$

where $\boldsymbol{x}_k$ ($k = 1, 2, \cdots, N$) represents an $M_x \times M_y \times M_z$-dimensional normalized vector of a subject in the dataset and $N$ is the number of subjects.

Note that in a specific group, the result for each subject was a matrix of $m \times m$, which not only reduced the complexity of classifications, but also projected the scatter in our preprocessed images to a new space for easy classifications.

#### 2.1.3 *PCA Based Method*

As PCA is a standard technique for extracting the most significant features[20], reducing the original data into a subset of features which contains the largest amount of variance, we also chose this method to perform dimension reduction. Here, we vectorized each normalized 3D image as a column vector in a high dimensional space, then constructed a new matrix which includes $N$ column vectors:

$$\boldsymbol{x} = [\boldsymbol{x}_1 \ \ \boldsymbol{x}_2 \ \ \cdots \ \ \boldsymbol{x}_N]. \quad (7)$$

The covariance matrix $\boldsymbol{M}_{\mathrm{C}}$ of the new dataset was then calculated as:

$$\boldsymbol{M}_{\mathrm{C}} = \frac{1}{N} \sum \boldsymbol{\varphi} \boldsymbol{\varphi}^{\mathrm{T}}, \quad (8)$$

$$\boldsymbol{\varphi} = \boldsymbol{x} - \boldsymbol{\mu}, \quad (9)$$

where $\boldsymbol{\mu}$ is the mean value of all samples.

Then, a set of eigenvectors corresponding to the brain covariance matrix are obtained. With the eigenvectors we can project image vectors onto a most distinguished space called eigenbrains space in which classification can be well performed.

#### 2.1.4 *Fisher Discriminant Ratio*

The selection of eigenvectors in LDA and PCA method will influence the final classification accuracy since discriminatory capacities of eigenvectors in different sets are diverse. Most studies tended to select the first 10 eigenvectors to compute the PCA projection axis. In this study, we use a criterion to choose $m$ leading eigenvectors according to their separation ability measured by FDR. The FDR is denoted as:

$$\mathrm{FDR} = \frac{(\mu_1^{\mathrm{FDR}} - \mu_2^{\mathrm{FDR}})^2}{\sigma_1^2 + \sigma_2^2}, \quad (10)$$

where $\mu_i^{\mathrm{FDR}}$ and $\sigma_i$ denote the within-class mean value and variance of the $i$th class for $i = 1$ and 2, respectively. With each possible value of $m$, we calculated the FDR of the two vectors, one for relative features and the other for non-relative features. Then the number of $m$ was chose to be the one that maximized the FDR.

### 2.2 Classification with SVM

The classification is achieved through SVM, a type of supervised machine learning, which has been shown to be of high accuracy with small training-and-testing errors. SVM separates a given set of binary labeled training data obtained by the aforementioned steps with a hyperplane (support vector) that is farthest from the two classes[21]. The hyperplane is defined as:

$$\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + \boldsymbol{w}_0 = \boldsymbol{0}, \quad (11)$$

where $\boldsymbol{w}$ is the weight vector and $\boldsymbol{w}_0$ is the offset which reflects the distance between the hyperplane and its nearest weight vector. Once the hyperplane is determined, new samples with unknown labels can be categorized according to the sign of the discriminant function:

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + \boldsymbol{w}_0. \quad (12)$$

When training data are too complexed to separate by a linear separation via SVM, kernel techniques are usually combined to make SVM work more effectively. We chose two types of kernels in the classiffier for our study.

The linear kernel is defined as

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j. \tag{13}$$

The decision function of the linear kernel can be expressed as:

$$\boldsymbol{f}(\boldsymbol{x}) = \mathrm{sign}\Big\{ \sum_{i=1}^{N_{\mathrm{s}}} \boldsymbol{\alpha}_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j \boldsymbol{y}_i + \boldsymbol{b} \Big\}, \tag{14}$$

where $N_{\mathrm{s}}$ is the support vector number, $\boldsymbol{y}_i$ is the corresponding label of training data, and $\boldsymbol{\alpha}_i$ is the Lagrange multipliers.

Another type is the radial basis function (RBF):

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathrm{e}^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}}, \tag{15}$$

where $\sigma$ is used to decide the width of the kernel function.

The corresponding decision function is:

$$\boldsymbol{f}(\boldsymbol{x}) = \mathrm{sign}\Big( \sum_{i=1}^{N_{\mathrm{s}}} \boldsymbol{\alpha}_i \boldsymbol{y}_i \mathrm{e}^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}} + \boldsymbol{b} \Big). \tag{16}$$

Note the input to the SVM classifier was the 22 AAL ROIs, selected either based on previous findings or via the feature selection process, the first $m$ most significant components of LDA or PCA. Here $m$ was different depending on groups.

## 3 Results

We performed the final classification via SVM with two different kernels for PCA and LDA based features, but for the ROI based features, only linear kernel is used due to the high dimension of the training features. Table 2 shows the final classification results of AD versus NC (Group 1), AD versus MCI (Group 2), and MCI versus NC (Group 3), where the accuracy (Acc), sensitivity (Sen) and specificity (Spe) are a series of statistical measurements of the performance of a binary classification test which are widely used to describe a diagnostic test, the positive/negative likelihood (PL/NL) is another assessment of the positive and predictive value of the study, given its prevalence independence and AUC is the area under the RDC curve. Among these indices, Acc is the rate of success. Acc, Sen and Spe are defined as:

$$\mathrm{Acc} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}), \tag{17}$$

$$\mathrm{Sen} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}), \tag{18}$$

$$\mathrm{Spe} = \mathrm{TN}/(\mathrm{FP} + \mathrm{TN}), \tag{19}$$

where TP, meaning true positive, and TN, meaning true negative, correspond to the numbers of patients and normal controls respectively which are correctly classified; FN means false negative and FP means false positive, refer to the numbers of subjects which fail to be classified as patients and normal controls respectively. Then expressions of PL and NL can be calculated as:

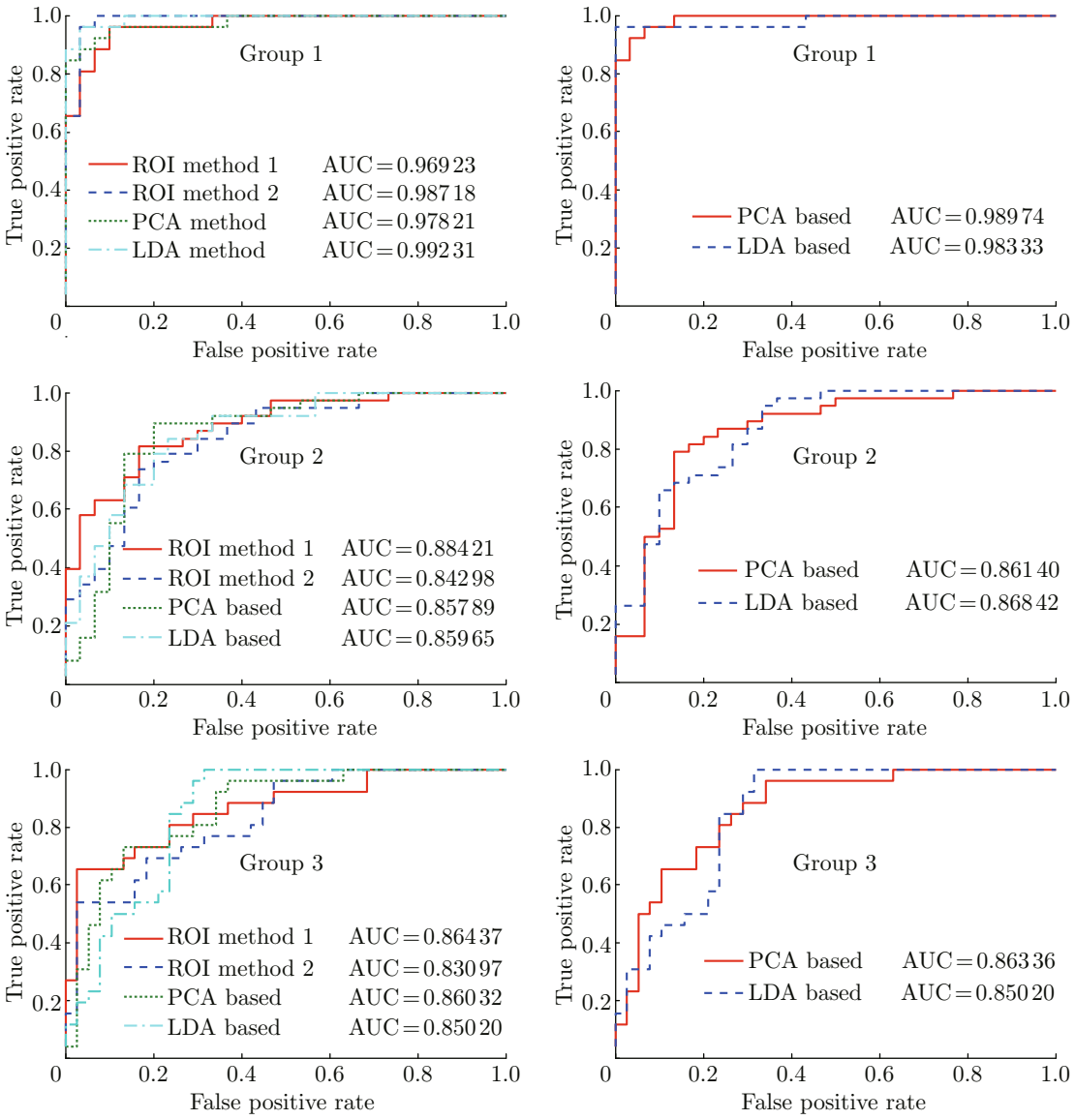
$$\mathrm{PL} = \mathrm{Sen}/(1 - \mathrm{Spe}) = \mathrm{TP}/\mathrm{TN}, \tag{20}$$

$$\mathrm{NL} = (1 - \mathrm{Sen})/\mathrm{Spe} = \mathrm{FN}/\mathrm{TN}. \tag{21}$$

Receiver operating curves (ROCs) of classification among all three groups are also obtained, as shown in Fig. 1, where images in the first column are results when

**Table 2 Comparison of methods used in our study**

| Method | | | Acc/% | Sen/% | Spe/% | PL | NL | AUC |
|---|---|---|---|---|---|---|---|---|
| AD versus NL | ROI | Method 1 | 89.29 | 93.33 | 84.62 | 6.067 | 0.079 | 0.962 |
| | | Method 2 | 91.10 | 96.70 | 84.60 | 6.290 | 0.040 | 0.987 |
| | PCA | Linear | 91.10 | 90.00 | 92.30 | 11.700 | 0.110 | 0.978 |
| | | RBF | 93.00 | 90.00 | 96.00 | 23.000 | 0.100 | 0.990 |
| | LDA | Linear | 94.60 | 96.70 | 92.30 | 12.600 | 0.040 | 0.992 |
| | | RBF | 92.90 | 98.90 | 90.60 | 10.500 | 0.010 | 0.983 |
| AD versus MCI | ROI | Method 1 | 79.41 | 70.00 | 86.84 | 5.319 | 0.346 | 0.884 |
| | | Method 2 | 75.00 | 70.00 | 79.00 | 3.330 | 0.380 | 0.843 |
| | PCA | Linear | 80.90 | 83.30 | 78.90 | 3.960 | 0.210 | 0.858 |
| | | RBF | 81.00 | 83.00 | 79.00 | 4.000 | 0.200 | 0.861 |
| | LDA | Linear | 79.40 | 76.70 | 81.60 | 4.160 | 0.290 | 0.860 |
| | | RBF | 77.90 | 78.90 | 73.10 | 2.930 | 0.290 | 0.868 |
| MCI versus NL | ROI | Method1 | 78.13 | 81.58 | 85.84 | 5.761 | 0.215 | 0.864 |
| | | Method 2 | 71.90 | 73.70 | 69.20 | 2.390 | 0.380 | 0.831 |
| | PCA | Linear | 75.00 | 71.10 | 80.80 | 3.650 | 0.360 | 0.860 |
| | | RBF | 77.00 | 71.00 | 85.00 | 4.600 | 0.300 | 0.863 |
| | LDA | Linear | 78.10 | 73.70 | 84.60 | 4.790 | 0.310 | 0.850 |
| | | RBF | 79.70 | 76.30 | 84.60 | 4.960 | 0.280 | 0.850 |




(a) Linear kernel (b) RBF kernel

Fig. 1 ROCs using the linear and RBF kernels

use the linear kernel, while those in the second column are with the RBF kernel.

Results of our method show that AD subjects can be distinguished from MCI or NC subjects at high accuracies of 81% and 94.6%, respectively. However, related pixel-based PET image analysis[22-23] achieved accuracies up to 74% and 88%, respectively. In addition, when using the improved LDA and PCA method, the results of AD versus MCI and MCI versus NC are both improved. The accuracy in distinguishing AD and NC is fairly high in both approaches, especially when using the linear kernel in SVM. However, it should be pointed out that, it is difficult to achieve a high accuracy for group 3 (MCI versus NC) due to the slight difference of brain metabolism among MCI and NC subjects. Our method showed an accuracy of 78.1% while the accuracy is up to 74% in the related studies[22-23].

## 4 Conclusion

With numerous reports in the AD literatures, AD researchers still feel the challenge not answered with regard to finding reliable and sensitive ways for early identification of the disease related abnormalities. We set up our study attempting to be a part of this big endeavor. We offered one approach primarily assembling various existing tools while in a form different from what has been reported in the literature, and evaluated the performance. The proposed approach, based on ROI, PCA and LDA methods for feature selections and support vector machines for the classification, demonstrated the effectiveness of classification with accuracies

up to 94.6% for AD versus NC, 81% for AD versus MCI and 78.1% for MCI versus NC.

The whole process consisted of processing images, selecting features (ROIs), and carrying out the SVM procedure. The processing of images, including image alignment, smoothing, and normalization, were conducted in SPM. The feature selection played an important role in dealing with high dimensional image data. If no feature selection was applied, the classification algorithm failed due to the huge computational load. More comprehensive feature selection methods are available and we chose the simple ones for the validation of the basic idea. Others could be relatively easy to examine given the established framework. In the classification, training with the SVM (without the construction of the support vector) was the most time consuming portion in the whole process. Fortunately the training could be done off-line before the whole process. It should also be noted that our current study is primarily for validating the basic idea and the whole procedure could be optimized further for future works, with more innovative methodology developments for computer-aided diagnosis of Alzheimer's disease.

**Acknowledgement** The authors thanks to Drs. Xiao Shi-fu and Wang Tao for their meticulous guidance and insightful comments.

## References

[1] Alzheimer's Association. 2013 Alzheimer's disease facts and figures [J]. *Alzheimer's & Dementia*, 2013, **9**(2): 208-245.

[2] Hampel H, Prvulovic D, Teipel S, et al. The future of Alzheimer's disease: The next 10 years [J]. *Progress in Neurobiology*, 2011, **95**(4): 718-728.

[3] Reiman E M, Jagust W J. Brain imaging in the study of Alzheimer's disease [J]. *NeuroImage*, 2012, **61**(2): 505-516.

[4] Querbes O, Aubry F, Pariente J, et al. Early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve [J]. *Brain*, 2009, **132**(8): 2036-2047.

[5] Duara R, Grady C, Haxby J, et al. Positron emission tomography in Alzheimer's disease [J]. *Neurology*, 1986, **36**(7): 879-887.

[6] Norderg A, Rinne J O, Kadir A, et al. The use of PET in Alzheimer disease [J]. *Nature Reviews Neurology*, 2010, **6**(2): 78-87.

[7] Jagust W J, Bandy D, Chen K, et al. The Alzheimer's disease neuroimaging initiative positron emission tomography core [J]. *Alzheimer's & Dementia*, 2010, **6**(3): 221-229.

[8] Foeter N L, Heidebrink J L, Clark C M, et al. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease [J]. *Brain*, 2007, **130**(10): 2616-2635.

[9] Du A-T, Schuff N, Kramer J H, et al. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia [J]. *Brain*, 2007, **130**(4): 1159-1166.

[10] Dickerson B C, Feczko E, Augustinack J C, et al. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area [J]. *Neurobiology of Aging*, 2009, **30**(3): 432-440.

[11] Gray K R, Wolz R, Keihaninejad S, et al. Regional analysis of FDG-PET for use in the classification of Alzheimer's disease [C]// *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro.* Chicago, USA: IEEE, 2011: 1082-1085.

[12] Gray K R, Woiz R, Heckemann R A, et al. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease [J]. *NeuroImage*, 2012, **60**(1): 221-229.

[13] Salmon E, Sadzot B, Maquet P, et al. Differential diagnosis of Alzheimer's disease with PET [J]. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, 1994, **35**(3): 391-398.

[14] Mosconi L. Glucose metabolism in normal aging and Alzheimer's disease: Methodological and physiological considerations for PET studies [J]. *Clinical and Translational Imaging*, 2013, **1**(4): 217-233.

[15] Illán I, Gorriz J, Lopez M, et al. Computer aided diagnosis of Alzheimer's disease using component based SVM [J]. *Applied Soft Computing*, 2011, **11**(2): 2376-2382.

[16] Kim E J, Cho S S, Jeong Y, et al. Glucose metabolism in early onset versus late onset Alzheimer's disease: An SPM analysis of 120 patients [J]. *Brain*, 2005, **128**(8): 1790-1801.

[17] Kono A K, Ishii K, Sofue K, et al. Fully automatic differential diagnosis system for dementia with Lewy bodies and Alzheimer's disease using FDG-PET and 3D-SSP [J]. *European Journal of Nuclear Medicine and Molecular Imaging*, 2007, **34**(9): 1490-1497.

[18] Reiman E, Chen K, Liu X, et al. Fibrillar amyloid-$\beta$ burden in cognitively normal people at 3 levels of genetic risk for Alzhimer's disease [J]. *Proceedings of the National Academy of Sciences*, 2009, **106**: 6820-6825.

[19] Noushath S, Hemantha K G, Shivakumara P. $(2D)^2$ LDA: An efficient approach for face recognition [J]. *Pattern Recognition*, 2006, **39**(7): 1396-1400.

[20] Zoua H, Hastiea T, Tibshirania R, et al. Sparse principal component analysis [J]. *Journal of Computational and Graphical Statistics*, 2006, **15**(2): 265-286.

[21] Wang L. Support vector machines: Theory and applications [M]. Berlin: Springer, 2005.

[22] Lopez M, Ramirez J, Gorriz J, et al. Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules [J]. *Electronics Letters*, 2009, **45**(8): 389-391.

[23] Ramirez J, Gorriz J, Segovia F, et al. Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification [J]. *Neuroscience Letters*, 2010, **472**(2): 99-103.