Computer Methods and Programs in Biomedicine 187 (2020) 105290

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Region-of-Interest based sparse feature learning method for Alzheimer's disease identification☆

Ling Wang[a], Yan Liu[b,*], Xiangzhu Zeng[c,**], Hong Cheng[a], Zheng Wang[c], Qiang Wang[d]

[a] Center for Robotics, University of Electronic Science and Technology of China, Chengdu, 611731 China
[b] School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, 100049 China
[c] Department of Radiology, Peking University Third Hospital, Beijing, 100191 China
[d] Beijing Union University, Beijing, 100101 China

ARTICLE INFO

Article history:
Received 2 September 2019
Revised 17 December 2019
Accepted 19 December 2019

Keywords:
Alzheimer's disease
Machine learning
Computer-aided disease diagnosis
Sparse feature learning
Elastic net

ABSTRACT

Background and Objective: In recent years, some clinical parameters, such as the volume of gray matter (GM) and cortical thickness, have been used as anatomical features to identify Alzheimer's disease (AD) from Healthy Controls (HC) in some feature-based machine learning methods. However, fewer image-based feature parameters have been proposed, which are equivalent to these clinical parameters, to describe the atrophy of regions-of-interest (ROIs) of the brain. In this study, we aim to extract effective image-based feature parameters to improve the diagnostic performance of AD with magnetic resonance imaging (MRI) data.

Methods: A new subspace-based sparse feature learning method is proposed, which builds a union-of-subspace representation model to realize feature extraction and disease identification. Specifically, the proposed method estimates feature dimensions reasonably, at the same time, it protects local features for the specified ROIs of the brain, and realizes image-based feature extraction and classification automatically instead of computing the volume of GM or cortical thickness preliminarily.

Results: Experimental results illustrate the effectiveness and robustness of the proposed method on feature extraction and classification, which are based on the sampled clinical dataset from Peking University Third Hospital of China and the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The extracted image-based feature parameters describe the atrophy of ROIs of the brain well as clinical parameters but show better performance in AD identification than clinical parameters. Based on them, the important ROIs for AD identification can be identified even for correlated variables.

Conclusion: The extracted features and the proposed identification parameters show high correlation with the volume of GM and the clinical mini-mental state examination (MMSE) score respectively. The proposed method will be useful in denoting the changes of cerebral pathology and cognitive function in AD patients.

1. Introduction

As a chronic neurodegenerative disease, Alzheimer's disease (AD) can be diagnosed from the change of cerebral cortex [1,2]. With the development of medical imaging technology, the images of cerebral cortex are widely used in AD clinical diagnosis. Recently, working as computer aided disease diagnosis technologies, many image processing methods and machine learning methods have been proposed to analyze magnetic resonance imaging (MRI) scanned data for early AD identification [3–20]. Useful information for diagnosis needs to be extracted from high dimension MRI data during machine learning. However, the problem of high dimensionality of MRI data with relatively small number of subjects is a widespread concern for most machine learning methods. Thus, many feature extractions based dimensionality reduction techniques have been applied.

Typically, there are two types of feature dimensionality reduction methods proposed for AD identification. One is linear-based subspace dimensionality reduction method [3–7], such as principal

☆ Research was supported in part by the "National Natural Science Foundation of China (NSFC)" (No. 61603077, No. 61305033, No. 41776204) and "Beijing Municipal Education Commission" (No. KM201611417014).

* Corresponding author.
** Co-corresponding author.
E-mail addresses: yanliu@ucas.ac.cn (Y. Liu), xiangzhuzeng@126.com (X. Zeng).

https://doi.org/10.1016/j.cmpb.2019.105290
0169-2607/

component analysis (PCA) [3], Fishers linear discriminant analysis (LDA) [4] and locality preserving projection (LPP) [21] combined with sparse analysis method [5]. Those high dimensional MRI data are projected into a low-dimensional subspace or a locally linear subspace to reduce the dimension of MRI data. PCA is an orthogonal transformation which convert observations samples of possibly correlated variables into linearly uncorrelated subspace called principal components (PCs). If there are n observations with p variables, then the number of distinct PCs is $\min(n-1, p)$. LDA is closely related to PCA in that they both look for linear combination of variables to best interpret the data. But LDA attempts to model the difference between classes data while PCA attempts to model the similarity within classes.

The other type of dimensionality reduction method is nonlinear technique, such as manifold learning methods [15–17] and multiple-kernel based methods [18,19]. These methods either provide an embedded mapping between high-dimensional space and low-dimensional space or give a distance measurement based on visualization. For example, manifold alignment is employed to learn joint low-dimensional manifold for 1.5T and 3T MR image intensity features, which are built by an adaptation of Laplacian eigen-maps [16]. Furthermore, statistical-based learning methods [22–25] and Topology-based methods [26] are also used for dimensionality reduction.

Considering that AD reflects anatomical atrophy or functional neurodegeneration in some regions-of-interest (ROIs) of brain, such as Amygdala, Hippocampus, Parahippocampal gyrus, Caudate, Fusiform gyrus, Thalamus, medial temporal lobe, etc [9], the change of their volumes and their localized shape deformations (i.e. the volume of gray matter and cortical thickness) have been used as anatomical features to distinguish AD from Healthy Controls (HC) [4,27] in most feature dimensionality reduction methods. However, these features are susceptible to the preprocessing results, and the preprocessing procedure may increase the complexity of feature learning. Meanwhile, for those PCA based methods, each principal component in PCA is a linear combination of all original variables. The number of PCs (i.e. feature dimension) is generally determined by cross-validation based methods, which may effect the identification results. Especially, for some ROIs with relative smaller volumes, some useful features are easier to lost if an unreasonable feature dimension is estimated. Recently, a local low-rank filtering method, called compartmental low-rank approximation (CLORA) method, is proposed for denoising practical magnetic resonance spectroscopic imaging (MRSI) data [28]. It performs local low-rank filtering compartmentally to protect weak and/or localized signals from being filtered out by global low-rank approximation.

In this paper, we proposed a compartmental sparse feature extraction and classification method for AD identification. Some preliminary accounts of this study were presented in our early conference papers [29,30]. The main contributions of this study are as follows:

(a) We build a union-of-subspace model with the assumption that each atlas-based ROIs of brain has similar neurodegeneration characteristics for AD. The proposed method partitions the T1-weighted MRI data into several compartments based on the local spatial support and estimates their corresponding feature dimensions using Singular Value Decomposition (SVD). It is worthwhile to note that local features with smaller volumes may be protected well based on this compartmental feature dimension estimation method.
(b) We perform compartmental sparse feature extraction by using sparse principal component analysis (SPCA) [31] method with the estimated feature dimensions to obtain sparse features for different ROIs. More specially, the extracted sparse principal components (SPCs) for each compartment reflect the volume changes of the specified ROIs of the brain.
(c) We propose a modified elastic net logistic regression (ENLR) method, called EN-ROC (i.e. EN combined with receiver operating characteristic curve) method, as classifier after dimensionality reduction. The identification parameters extracted from EN-ROC method show better performance in AD identification.

It is noted that even some deep learning classifiers have shown good performance on AD classification [32,33], it is difficult to present interpretability on the extracted features or classification results because they incorporate feature extraction and classifier learning into an unified framework that is typically regarded as black boxes [34]. Furthermore, in this work, we aim to extract effective image-based features, which are comparable with those clinical parameters, to describe the atrophy of ROIs of the cerebrum. Because of these reasons, deep learning classifiers will not be discussed in this work.

2. Method

The framework of the proposed method is shown in Fig. 1. First, image preprocessing is used to perform segmentation on the original MRI T1-weighted images. Second, based on the derived ROIs masks, compartmental model is built to estimate feature dimensions of different ROIs. Third, the compartmental sparse feature extraction with estimated feature dimensions is performed on different ROIs to extract features and corresponding projection vectors. Finally, the test samples are projected into feature space and identified based on EN-based classifiers.

2.1. Materials and image preprocessing

In this study, experimental data includes two parts: dataset $\{\Phi_1\}$ is sampled from Peking University Third Hospital of China, and dataset $\{\Phi_2\}$ is downloaded from Alzheimer's Disease Neuroimaging Initiative (ADNI) website (http://www.loni.ucla.edu/ADNI). ADNI website was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations. Dataset $\{\Phi_1\}$ includes 131 AD subjects and 131 HC subjects, which are sampled on a SIEMENS Trio 3T scanner with the acquisition parameters $TR/TE = 2350/3.44ms$, voxel size $= 1 \times 1 \times 1mm^3$. These data is approved by the local Institutional Review Board. And dataset $\{\Phi_2\}$ includes 67 AD subjects and 105 HC subjects, which are sampled from different 3T scanners (i.e. GE, SIEMENS, PHILIP) and are used to illustrate the effectiveness and robustness of the proposed method. Both of these two datasets are all T1 Magnetization Prepared Rapid Gradient Echo (MPRAGE) data.

1) Subjects

The selected subjects follow the general inclusion/exclusion criteria as described in Table 1, i.e. the MMSE score of each HC

Table 1
Demographics of the subjects.

	AD	HC
Size	198	236
Female/Male	110/88	120/116
Age	72.61 ± 8.15	71.58 ± 7.32
Education	11.1 ± 5.78	14.18 ± 3.34
[a]MMSE	17.84 ± 6.11	29.06 ± 1.02
[b]CDR	0.5, 1	0

[a] MMSE: mini-mental state examination.
[b] CDR: clinical dementia rating.

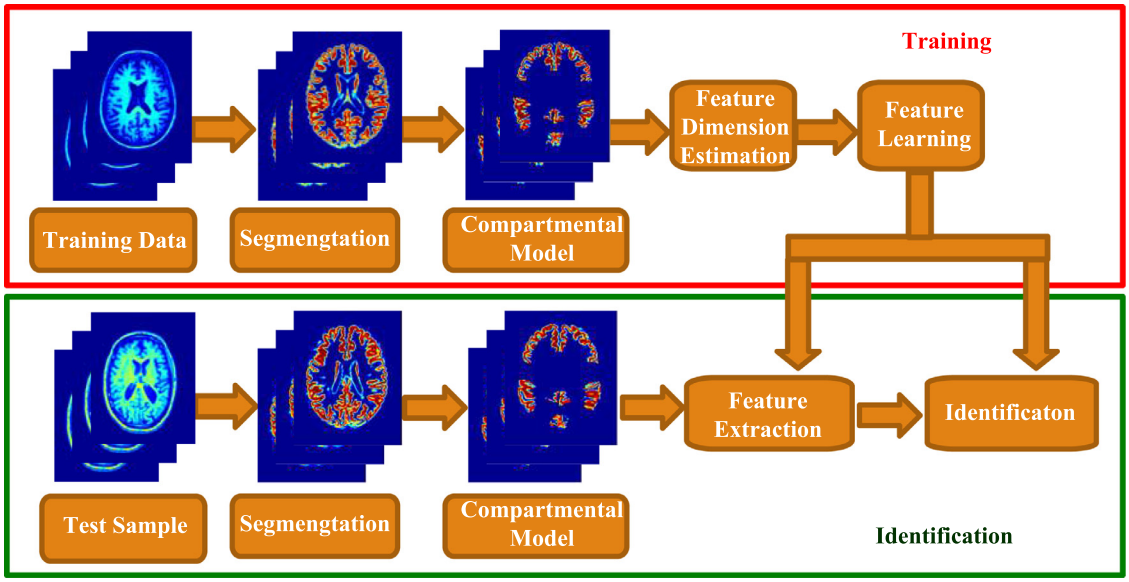


Fig. 1. The Framework of the proposed method.

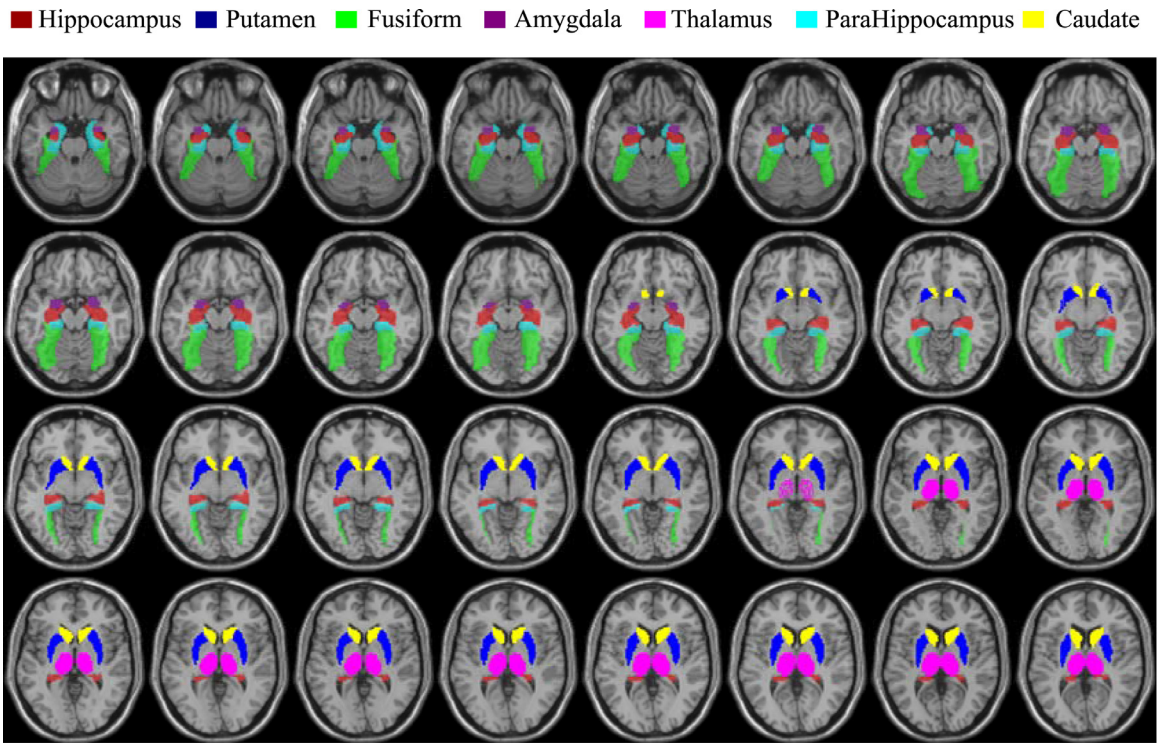


Fig. 2. Seven ROIs masks in a slice view.

subject is between 26 and 30 with clinical dementia rating (CDR) of 0. And the MMSE score of each AD subject is between 10 and 24 with the CDR of 0.5 or 1.

2) Image preprocessing

All datasets are reviewed for quality and further preprocessed with SPM8 [35] that is running under a Windows MATLAB platform by the following steps.

First, all T1-weighted images are segmented into GM, white matter (WM) and cerebrospinal fluid (CSF) maps in the 'realigned' and 'warped' template by using the 'new segment' toolbox in SPM8.

Second, 45 ROIs masks are derived by using WFU PickAtlas [36] based on the automated anatomical labeling (AAL) [37] template using GM data. In this study, both left and right structures of cerebrum in AAL template are combined into one to reduce the number of ROIs from 90 to 45. Each mask is used as the spatial support of indicator function $\mathbf{W}_q(\mathbf{r})$ for its corresponding compartment as described in (1). For example, seven ROIs masks are shown in a slice view in Fig. 2, which are used as spatial support for seven corresponding indicator functions. These ROIs are selected to show experimental results in Section 3.

Then the following feature dimension estimation is performed based on these simply preprocessed data.

2.2. Proposed method

1) Compartmental feature dimension estimation

In this study, we use CLORA model [28] as an union-of-subspace representation for the ROIs of the brain. There are two fundamental assumptions in CLORA model. First, we assume that the measured MPRAGE brain data can be partitioned into several compartments, in which each compartment corresponds to one atlas-based ROI. Second, we assume that each compartment has similar neurodegeneration characteristics for AD. This assumption is valid because each atlas-based ROIs generally reflects similar anatomical atrophy or functional neurodegeneration, e.g. Hippocampus is mainly associated with memory. Under these assumptions, the sampled MRI brain data can be partitioned into different compartments and presented as the following compartmental model:

$$\mathbf{S}(\mathbf{r}) = \sum_{q=1}^{Q} \mathbf{W}_q(\mathbf{r}) \odot \mathbf{S}^{(q)}(\mathbf{r}), \tag{1}$$

where $\odot$ denotes matrix dot product, $\mathbf{r}$ denotes the 3-D spatial coordinates, $\mathbf{S}(\mathbf{r})$ is the measured MRI data of the whole brain or specified brain region, Q is the type of different ROIs (i.e. number of compartments), $\mathbf{W}_q(\mathbf{r})$ is the corresponding indicator function of $\mathbf{S}^{(q)}(\mathbf{r})$ and can be derived from 45 ROIs masks in image preprocessing procedure (i.e. $\mathbf{W}_q(\mathbf{r}) = 1$ for $\mathbf{r} \in \mathcal{D}_q$ and $\mathbf{W}_q(\mathbf{r}) = 0$ for others, where $\mathcal{D}_q$ represents the spatial support of $\mathbf{S}^{(q)}(\mathbf{r})$), $\mathbf{S}^{(q)}(\mathbf{r})$ is the q-th compartment and can be approximated using SVD as

$$\mathbf{S}^{(q)} = \sum_{l=1}^{k_q} \sigma_l^{(q)} u_l^{(q)} v_l^{(q)T}, (q = 1, 2, \cdots, Q), \tag{2}$$

where $\sigma_l^{(q)}$, $u_l^{(q)}$ and $v_l^{(q)}$ are the singular value, left singular vector, and right singular vector of $\mathbf{S}^{(q)}$, respectively.

Based on this compartmental model, we use k_q as the estimated feature dimension for the q-th compartment, i.e. the number of PCs of $\mathbf{S}^{(q)}$ in SPCA method. In order to obtain a reasonable feature dimension while protecting local features, we maximize k_q for each compartment to maintain most features of $\mathbf{S}^{(q)}$, i.e. if $\sigma_l > D_q$ and $\sigma_{l+1} < D_q$, where D_q is the threshold value of the singular value distributions for the qth compartment. Specifically, for those specified brain regions with several ROIs, most local features with smaller volumes may be protected well by using this compartmental feature dimension estimation method. This point will be illustrated in Section 4 as shown in Fig. 6. In this study, AHP (Amygdala+Hippocampus+Parahippocampal gyrus) is a combined specified brain region, in which local features for each independent ROIs (i.e. Amygdala, Hippocampus, or Parahippocampal gyrus) have lower dimensionality. As Fig. 6 shows, the classification accuracy for AHP or each independent ROIs based on the proposed method is higher than that of unreasonable preset fixed values of $k = 5, 10$ and is comparable with that of other higher fixed values of $k = 20, 50, 80$. Thus, we may analyze the effectiveness of a single ROI or the combined ROIs on AD progression by using the estimated feature dimension and SPCA method on different compartments.

In this study, we obtain 45 individual compartments by masking derived ROIs on the segmented GM data separately. It is necessary for each compartment to be preprocessed further for the following feature extraction. As discussed in [29], in order to reduce the high dimension of 3D MRI images, all compartment are summed along the direction of axial to obtain 2-D compartmental data. Then the 2-D compartmental data are reshaped to a row vector $\mathbf{x}^{(q)}$. For all measured n samples, the q-th compartmental observation matrix $\mathbf{X}^{(q)}$ for the following feature extraction can be built as $\mathbf{X}^{(q)} = [\mathbf{x}^{1(q)}; \cdots; \mathbf{x}^{i(q)}; \cdots; \mathbf{x}^{n(q)}]$, $(i = 1, \cdots, n)$. For simplicity, we use $\mathbf{X}$ denotes $\mathbf{X}^{(q)}$ in the rest of this paper. In this study, we mainly aim to do 2D feature extraction, 3D feature extraction work will be discussed in the future work.

2) Sparse feature extraction

As we know, PCA is a low-rank approximation method, which is aiming to find linearly uncorrelated orthogonal basis and PCs to reduce the dimension of matrix. It usually can be computed according to SVD or Eigen-Decomposition (ED). Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be centered, where n and p are the numbers of samples and variables separately. Let the SVD of $\mathbf{X}$ be $\mathbf{U}_0\mathbf{D}_0\mathbf{V}_0^T$. Then we can obtain the linearly uncorrelated variables (i.e. PCs) of $\mathbf{X}$ in the subspace as $\mathbf{Z}_0 = \mathbf{U}_0\mathbf{D}_0 = \mathbf{X}\mathbf{V}_0$. The columns of $\mathbf{U}_0$ are called the left-singular vectors of $\mathbf{X}$, and the columns of $\mathbf{V}_0$ are the corresponding projection operators of PCs. The diagonal elements of $\mathbf{D}_0$ are corresponding sample variances of PCs. Usually, the first $k(k \ll \min(n, p))$ PCs which corresponding to the biggest k variances are chosen to represent the matrix $\mathbf{X}$, thus the dimensionality reduction is achieved. For the orthogonality of $\mathbf{U}_0$, $\mathbf{D}_0$ and $\mathbf{V}_0$, the PCs are uncorrelated to each other and the elements of $\mathbf{V}_0$ are typically nonzero. Each PC is a linear combination of all p variables and the projection vectors are typically nonzero, which often makes PCA difficult to interpret the derived PCs [31]. For feature extraction, PCA only selects at most n variables if $n < p$. And if there is a group of variables which are correlated with each other, PCA only selects one variable randomly from the group.

To overcome those drawbacks of PCA, Zou *et al.* proposed SPCA method [31], where the PCA was formulated as a regression-type optimization problem by combining ridge regression and least absolute shrinkage and selection operator (LASSO) regression [38]. SPCA is an EN-based [39] sparse feature extracting method, which can be formulated as the following two-step iterative optimization problem:

$$\min_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{n} \|\mathbf{x}^i - \mathbf{A}\mathbf{B}^T\mathbf{x}^i\|^2 + \lambda_1 \sum_{j=1}^{k} \|\beta_j\|^2 + \sum_{j=1}^{k} \lambda_{2,j}\|\beta_j\|_1,$$
$$subj.to \qquad \mathbf{A}^T\mathbf{A} = \mathbf{I}_{k\times k}, \tag{3}$$

where $\mathbf{x}^i$ $(i = 1, \cdots, n)$ is the vector of the i-th row of matrix $\mathbf{X}$, $\mathbf{A}$ is a transition matrix, $\mathbf{B} = [\beta_1, \cdots, \beta_k]$, $\beta_j \in \mathbb{R}^p$ $(j = 1, \cdots, k)$, is the sparse projection matrix of PCs, k is the number of selected PCs, λ_1 and $\lambda_{2,j}$ are non-negative Lagrangian penalizing parameters. λ_1 is used to overcome the over-fitting problem, and $\lambda_{2,j}$ is used to control the sparsity of β_j, i.e. the larger $\lambda_{2,j}$ the more sparse β_j. $\|\cdot\|^2$ and $\|\cdot\|_1$ are ℓ_2 and ℓ_1 norm of a vector respectively. Then SPCs of $\mathbf{X}$ is

$$\mathbf{Z} = \mathbf{X}\mathbf{B}. \tag{4}$$

The sparse projection matrix $\mathbf{B}$ can be obtained by solving (3) iteratively as described in [31]:

(a) Let $\mathbf{A} = [\alpha_1, \cdots, \alpha_k]$, $\alpha_j \in \mathbb{R}^p$, start as the project loadings of $\mathbf{X}$ (i.e. the first k of $\mathbf{V}_0$) or a random matrix.

(b) Given a fixed $\mathbf{A}$, the optimal $\mathbf{B}$ can be obtained by solving the following EN regression problem,

$$\arg\min_{\mathbf{B}} \sum_{j=1}^{k} \|\mathbf{X}\alpha_j - \mathbf{X}\beta_j\|^2 + \lambda_1\|\beta_j\|^2 + \lambda_{2,j}\|\beta_j\|_1. \tag{5}$$

(c) For a fixed $\mathbf{B}$, the penalty part in (3) can be ignored,

$$\min_{\mathbf{A}} \sum_{i=1}^{n} \|\mathbf{x}^i - \mathbf{A}\mathbf{B}^T\mathbf{x}^i\|^2 = \|\mathbf{X} - \mathbf{X}\mathbf{B}\mathbf{A}^T\|^2,$$
$$subj.to \qquad \mathbf{A}^T\mathbf{A} = \mathbf{I}_{k\times k}. \tag{6}$$

Then the solution of (6) can be obtained by computing the SVD of $\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, and update $\mathbf{A} = \mathbf{U}\mathbf{V}^T$.

(d) Repeat the step (a) and the step (b) until convergence.

Thanks to EN regression technique, the following factors help SPCA overcome limitations of PCA. (1) SPCA method will potentially select all p variables as features when it is applied to high dimension and relative low sample size data, i.e. $p \gg n$. (2) SPCA method will identify important variables while eliminate trivial ones based on the derived SPCs. (3) SPCA method is able to select group-variables due to the grouping effects of EN [39], i.e. SPCA is able to identify important variables with high explained variance even for correlated variables. The grouping effects of EN mean that the absolute value of regression coefficients of a group of highly correlated variables tend to be equal, which makes all correlated variables can be selected.

Based on these factors, SPCA method can extract important variables correctly when there are correlated variables, which makes the interpretation of the derived SPCs become easier [31]. These points will be illustrated in Fig. 9 in Section 4.

3) Sparse classification

After feature extraction, classifier is used for disease identification, such as SVM [3,40,41], EN [39] and LASSO [42]. Considering the comparison results discussed in [43], we use EN based method as classifier on the dimension reduced data.

Based on the estimated feature dimensions, we can obtain SPCs of the training data as $\mathbf{Z} = \mathbf{XB}$. For a test sample $\mathbf{y} \in \mathbb{R}^p$ of the c-th class, its corresponding SPCs can be represented as $\mathbf{y}_{te} = \mathbf{B}^T\mathbf{y}$, which is used as the extracted sparse feature for the following identification.

Let C be the number of class, n_c $(c = 1, \ldots, C)$ be the number of samples in the c-th class, $\sum_{c=1}^{C} n_c = n$, and ς_c is the label of the c-th class. In this study, based on ENLR method [39], we built the following three different classifiers to identify AD from HC: EN-ORI(i.e. original ENLR) [39], EN-SPC (i.e. EN combined with sparse principle components) [29], and EN-ROC [44].

(a) EN-ORI

In the statistical learning theory, the classification problem can be formulated as a linear regression model. Considering ENLR is a regression model [39], we use it as a classifier directly. Given $\mathbf{Z}$ and their label $\mathbf{v} = [v_1, \cdots, v_n]^T$ $(v_i \in \{\varsigma_c\}_{c=1}^{C})$. Without loss of generality, $\mathbf{v}$ is centered and $\mathbf{Z}$ is standardized. The identification can be performed according to the following rules:

$$\hat{c} = \arg\min_{c} \|\mathbf{y}_{te}^T\hat{\eta} - \varsigma_c\|^2, \tag{7}$$

where $\hat{\eta}$ is the projection vector of SPCs,

$$\hat{\eta} = \arg\min_{\eta} \|\mathbf{v} - \mathbf{Z}\eta\|^2 + \mu_1\|\eta\|^2 + \mu_2\|\eta\|_1, \tag{8}$$

where μ_1 and μ_2 are non-negative Lagrangian parameters the same as (3).

(b) EN-SPC

Unlike EN-ORI, EN-SPC formulates the regression representation based on the sparse linear combination of training samples rather than training labels [29], i.e., the test sample is represented by an over-complete dictionary $\mathbf{Z}$ with sparse basis. Considering the noises and interferences, the classifier can be expressed as follows:

$$\hat{c} = \arg\min_{c} \|\mathbf{y}_{te} - \mathbf{Z}^T\phi_c(\hat{\gamma})\|^2, \tag{9}$$

where $\hat{\gamma} = [\hat{\gamma}_{1,1}, \cdots, \hat{\gamma}_{1,n_1}, \cdots, \hat{\gamma}_{c,1}, \cdots, \hat{\gamma}_{c,n_c}, \cdots, \hat{\gamma}_{C,1}, \cdots \hat{\gamma}_{C,n_C}]^T$ is a vector of linear regression coefficients, $\phi_c(\hat{\gamma}) \in \mathbb{R}^n$ includes nonzero items of $\hat{\gamma}$ that are associated with c-th class, and $\phi_c(\hat{\gamma}) = [0, 0, \cdots, \hat{\gamma}_{c,1}, \cdots, \hat{\gamma}_{c,n_c}, 0, \cdots, 0]^T$ is a sparse basis vector.

$$\hat{\gamma} = \min_{\gamma} \|\mathbf{y}_{te} - \mathbf{Z}^T\gamma\|^2 + \lambda_3\|\gamma\|^2 + \lambda_4\|\gamma\|_1, \tag{10}$$

where λ_3 and λ_4 are non-negative Lagrangian parameters, which are chosen based on the discrepancy principle although many other regularization parameter selection schemes can be used. Classification results with the effect of different λ_4 are shown in Section 3.

For $\mathbf{Z} \in \mathbb{R}^{n\times k}$ with $k \ll n$, EN-SPC classifier is an ENLR based sparse classification method, which is different from LASSO expression in [42].

(c) EN-ROC

The ROC analysis provides a method to select possibly optimal models, in which the classifier boundary between classes is determined by a threshold value. ROC curve describes a relationship between the true positive rate (TPR) and the false positive rate (FPR) at various threshold settings, which is used to illustrate the diagnostic ability of a binary classifier system. Based on (10), we modify the EN-SPC classifier by choosing an optimal threshold value in ROC curve [44]. Define an identification ratio as

$$\rho_c = \|\mathbf{Z}^T\phi_c(\hat{\gamma})\|^2/\|\mathbf{Z}^T\hat{\gamma}\|^2, \tag{11}$$

where the description of $\phi_c(\hat{\gamma}) \in \mathbb{R}^n$ is shown in (9). Then the modified identification rule can be described as

$$\hat{c} = \begin{cases} \varsigma_{\mathrm{p}} & \text{if} \rho_c > \hat{\delta}_c \\ \varsigma_{\mathrm{n}} & \text{otherwise} \end{cases}, \tag{12}$$

where ς_{p} and ς_{n} are the labels of positive instances and negative instances respectively. If the discrimination threshold value (i.e. the identification ratio ρ_c) is varied from 0 to 1 continuously, we can obtain a ROC curve to illustrate the diagnostic ability of the classifier. And the optimal threshold value $\hat{\delta}_c$ in ROC curve be defined as the maximum difference between TPR and FPR,

$$\hat{\delta}_c = \arg\max_{\delta_c}(\mathrm{TPR}_{\delta_c} - \mathrm{FPR}_{\delta_c}). \tag{13}$$

In this study, AD is set as a positive instance and HC is set as a negative instance. Then $\phi_c(\hat{\gamma})$ in (11) is associated with AD.

4) Summary of the proposed method

The proposed method performs feature dimension estimation, feature extraction and classification compartmentally based on a union-of-subspace model representation. The data processing steps involved are summarized as follows.

(a) Derive atlas-based spatial masks $\mathbf{W}_q(r)$ in (1) by using new segment toolbox in SPM8 and WFU_PickAtlas based on AAL template.

(b) Perform compartmental low-rank approximation by using SVD to estimate feature dimensions for each compartment $\mathbf{S}^{(q)}$ in (2).

(c) Build the observation matrix $\mathbf{X}$ using all row vectors $\mathbf{x}^{i(q)}$, $(i = 1, \cdots, n)$ of the qth compartment.

(d) Perform sparse feature extraction for each compartmental data by solving the optimization problem in (3). More specifically, for $p \gg n$, (5) can be solved by a fast soft-thresholding algorithm as the following [31]:

$$\beta_j^{(q)} = \left(|\xi| - \frac{\lambda_{2,j}}{2}\right)_{+} \mathrm{Sign}(\xi), \tag{14}$$

where $\xi = \alpha_j^T\mathbf{X}^{(q)T}\mathbf{X}^{(q)}$, $(x)_+$ denotes $\max(0, x)$, Sign($\cdot$) is the sign function.

(e) Obtain the identification results by using three different classifiers that have been described in 3) of Section 2.2.

3. Results

In this study, in order to avoid any possible bias during samples partitioning, 50 times Monte-Carlo simulations were carried out to:

(a) estimate compartmental feature dimension k and select Lagrangian parameters $\lambda_{2,j}$ and λ_4 in (14) and (10),
(b) demonstrate the performance of three feature extraction methods (i.e. PCA, LDA, SPCA),
(c) demonstrate the performance of four classifiers (i.e. SVM, EN-ORI, EN-SPC and EN-ROC).

The final result was the average of the results of 50 times. 155 AD subjects and 155 HC subjects were randomly selected for training while remaining for testing during each Monte-Carlo simulation. In this study, 45 derived ROIs and 2 combined ROIs (i.e. AHP and whole cerebral GM) were all trained and tested. Considering the following previous studies [45–47], we selected 9 ROIs that are associated with memory function to show experimental results, i.e. Amygdala (Amgy), Caudate (Caud), Fusiform (Fusif), Hippocampus (Hippo), Parahippocampal gyrus (ParaH), Putamen (Put), Thalamus (Thal), AHP and whole cerebral GM (All). According to the diagnostic experience and previous studies [9], we hypothesized that Hippocampus, Parahippocampal gyrus and medial temporal lobe could achieve the best performance.

First, we selected the feature dimension k in (3) for each compartment. In this work, we chose k_q in (2) as the estimated feature dimension with the threshold value $D_q = 1e-10$ for the q-th compartment. Thus, the selected feature dimension k_q for the specified ROIs were marked as shown in Fig. 3.

Then, 50 times Monte-Carlo experiment were used to address the values of two Lagrangian penalizing parameters $\lambda_{2,j}$in (14) and λ_4 in (10). In this study, for $p \gg n$, (5) could be solved by a fast soft-thresholding algorithm as described in (14), which only needed to select Lagrangian penalizing parameters λ_2. The solution of (10) was as same as (5). Thus we only discussed the selection of $\lambda_{2,j}$ and λ_4 in this section. As described in Section 2, $\lambda_{2,j}$ effects the sparsity of the projection vector β_j in (14), λ_4 effects the sparsity of the projection vector γ in (10). β_j and γ became more sparse with the larger $\lambda_{2,j}$ and λ_4. In order to reduce complexity, $\lambda_{2,j}$ was defined to be a unified value as λ_2. The performances of different λ_2 and λ_4 are shown in Fig. 4. As seen from Fig. 4, classification accuracy is not influenced too much by different λ_2 and λ_4. As a result, we chose $\lambda_2 = 50$ and $\lambda_4 = 50$.

Table 2
Correlation between different parameters of hippocampus*.

	Volume of GM	MMSE score	$\|\|y_{te}\|\|^2$	ρ_c
Volume of GM	1	0.393	**0.936**	0.5
MMSE score	0.393	1	0.586	**0.776**
$\|\|y_{te}\|\|^2$	**0.936**	0.586	1	0.718
ρ_c	0.5	**0.776**	0.718	1

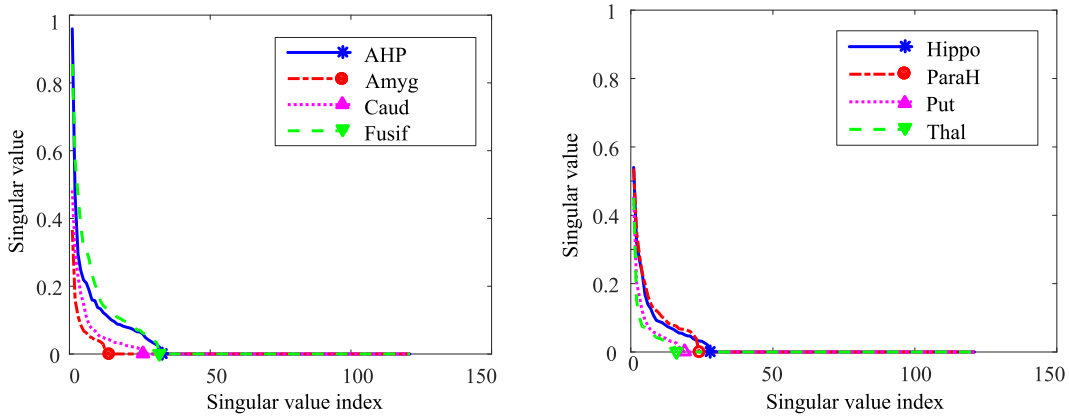
*$p < .001$.

Table 3
Classification results of different classifiers–Accuracy. (Boldface denotes the highest classification ratio in each ROIs.).

ROIs	Accuracy				
	SVM-G	SVM-L	EN-ORI	EN-SPC	EN-ROC
Amyg	0.731	**0.750**	0.711	0.702	0.739
Hippo	0.844	0.854	0.833	0.875	**0.878**
ParaH	0.834	0.852	0.816	0.843	**0.868**
AHP	0.841	0.860	0.827	0.868	**0.887**
Caud	0.732	0.718	0.720	0.717	**0.747**
Put	0.717	0.682	0.664	0.694	**0.705**
Thal	0.707	**0.734**	0.668	0.713	0.727
Fusif	0.604	0.638	0.589	**0.664**	0.656
All	0.846	0.836	0.800	0.831	**0.864**

In order to validate the performance of the extracted features (i.e. the extracted SPCs) and the proposed classifiers, here we randomly selected 50 AD subjects and 50 HC subjects as testing samples, and the rest as training samples. Correlation coefficients between different parameters (volume of GM, clinical MMSE score, L_2 norm of SPCs of test samples $||y_{te}||^2$ and identification ratio ρ_c) of Hippocampus are shown in Table 2. It is noticeable from Table 2 that $||y_{te}||^2$ shows high correlation with the volume of GM, which implies that the extracted SPCs for this compartment reflect the volume changes of the Hippocampus. Furthermore, from Fig. 5, we can see that $||y_{te}||^2$ has shown better performance than the volume of GM in AD identification. Thus, it is reasonable to use $||y_{te}||^2$ as sparse features in AD identification.

Finally, based on the extracted features, the classification performance between SVM-based classifiers and EN-based classifiers (i.e. EN-ORI, EN-SPC and EN-ROC) were compared in Tables 3–5. In this work, for SVM-based classifers, Gaussian kernel (i.e. SVM-G) and Linear kernel (i.e. SVM-L) functions were used to compute Gram matrix in SVM training procedure separately. As seen from Tables 3–5, EN-ROC classifier shows the highest classification



(a) (b)

Fig. 3. The selection of k. The vertical axis is the normalized singular value. The horizontal axis is the singular value index of the specified ROIs for different compartments and the selected rank k_q is marked by '*', '●', '▲', '▼' respectively.

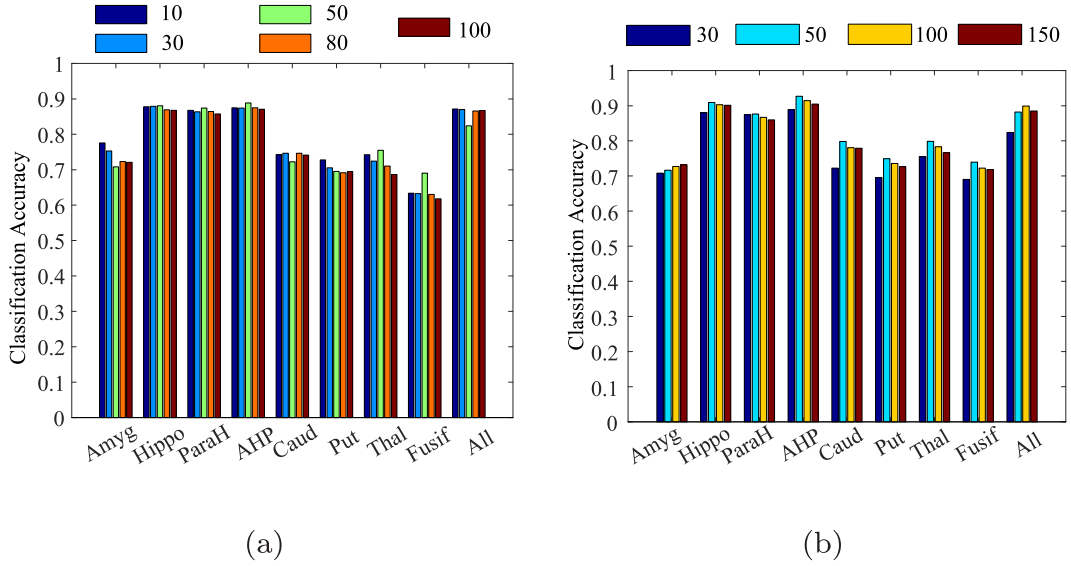


Fig. 4. Performance comparison with different λ_2 and λ_4. (a) Classification accuracy of using different λ_2. (b) Classification accuracy of using different λ_4.

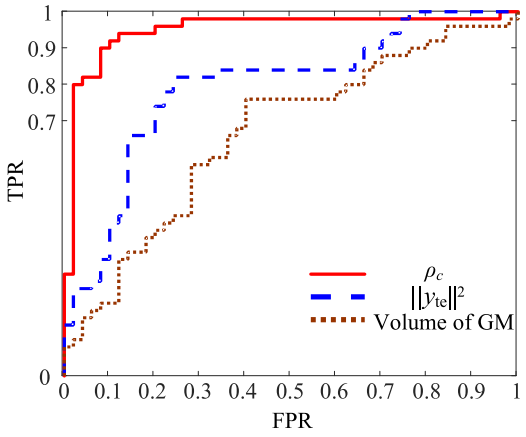


Fig. 5. ROC curve of different parameters for Hippocampus. The AUC for the volume of GM is 0.667, and for the extracted SPCs $||y_{te}||^2$ is 0.792, and for the identification ratio ρ_c is 0.952.

Table 4
Classification results of different classifiers–TPR. (Boldface denotes the highest classification ratio in each ROIs.).

ROIs	TPR				
	SVM-G	SVM-L	EN-ORI	EN-SPC	EN-ROC
Amyg	0.742	0.746	**0.778**	0.690	0.707
Hippo	0.849	0.863	0.884	0.864	**0.929**
ParaH	0.811	0.819	**0.867**	0.834	0.847
AHP	0.849	0.863	**0.910**	0.868	0.880
Caud	0.747	0.766	**0.820**	0.747	0.791
Put	0.680	0.690	0.687	0.651	**0.748**
Thal	0.756	0.710	0.693	0.731	**0.786**
Fusif	0.642	**0.790**	0.639	0.592	0.741
All	0.818	0.850	**0.892**	0.874	0.857

accuracy in most of ROIs, SVM-L, SVM-G and EN-SPC show better performance than EN-ORI, and SVM-L shows better performance than SVM-G for the SVM-based classifiers.

With TPR and FPR be considered, EN-ROC provides better balance between classification accuracy, TPR and FPR than other three classifiers. Even though the EN-ORI gives the most highest TPRs, those FPRs are higher too. The main reason is that the projection vector η in (8) is a relatively rough estimation for class labels while

Table 5
Classification results of different classifiers–FPR. (Boldface denotes the highest classification ratio in each ROIs.).

ROIs	1-FPR				
	SVM-G	SVM-L	EN-ORI	EN-SPC	EN-ROC
Amyg	0.725	0.753	0.674	0.708	**0.757**
Hippo	0.841	0.848	0.805	**0.880**	0.850
ParaH	0.834	0.871	0.788	0.848	**0.886**
AHP	0.837	0.858	0.780	0.868	**0.891**
Caud	0.692	0.725	0.665	0.700	**0.723**
Put	0.678	0.737	0.651	**0.710**	0.681
Thal	0.680	**0.747**	0.654	0.703	0.695
Fusif	0.583	0.553	0.561	**0.703**	0.609
All	0.862	0.828	0.749	0.807	**0.867**

test sample in (10) is represented by the SPCs of training samples. Considering these results, we chose EN-ROC as classifier after feature extraction in the following experiments.

From Table 2 and Fig. 5, we can see that the identification ratio ρ_c proposed in EN-ROC classifier shows higher correlation with clinical MMSE parameter and better performance in AD identification than traditional parameter, i.e. volume of GM. It means that the proposed compartmental sparse feature extraction method and EN-ROC classifier are effective and robust in identification between AD and HC with the dataset from different scanners.

4. Discussion

First, we investigated the effects of feature dimension on classification accuracy. The proposed method has been implemented in MATLAB R2014a, running on a computer with Intel Quad Core i5 at 3.4 GHz and 8 GB of RAM. From Fig. 6, we can see that the proposed compartmental feature dimension estimation method provides comparable classification performance compared with preset fixed values of $k = 50, 80,$ but their computing time is much longer than the proposed method. Considering the classification accuracy and computing time, we can see that the proposed method estimates the reasonable feature dimensions compartmentally for the following feature extraction and provides a higher computational efficiency for classification.

Secondly, we investigated the convergence of SPCA. SPCA was used to extract the SPCs on the Hippocampus image data, where

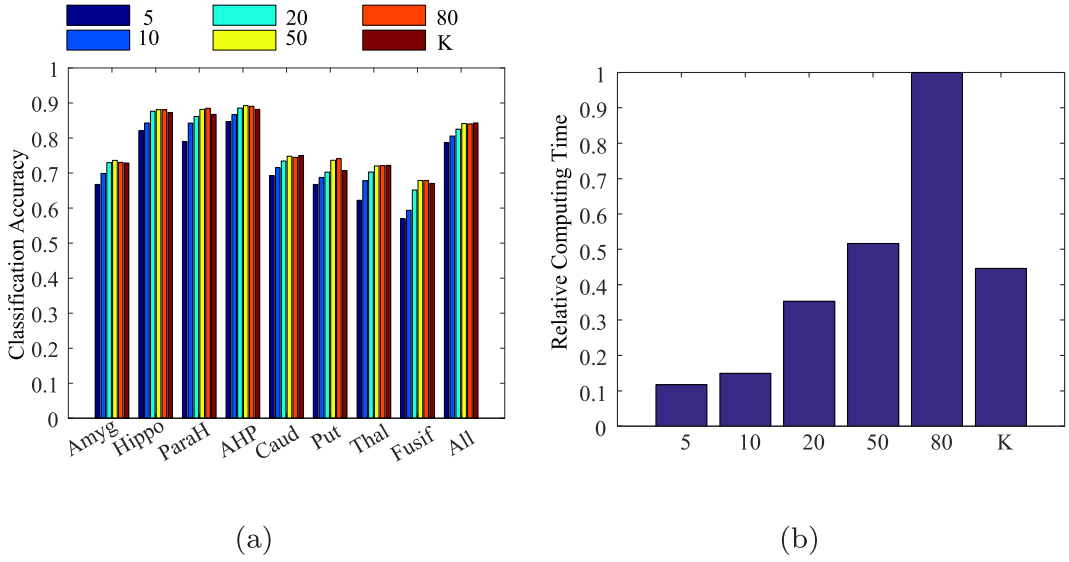


Fig. 6. Performance comparison with different feature dimensions. (a) Comparison of classification accuracy for different ROIs with different feature dimensions. (b) Relative computing time with different feature dimensions. Here, the longest training time is nearly 2 minutes/ROI for 310 samples.

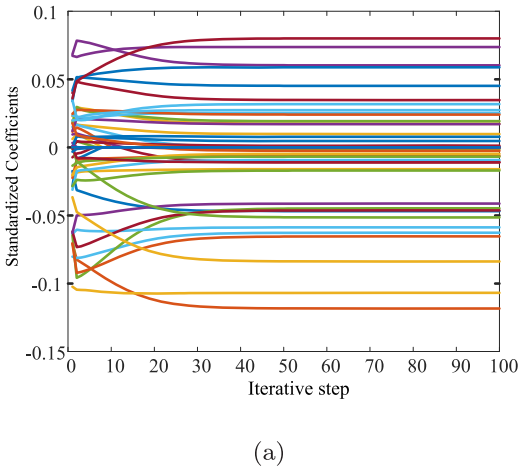


Fig. 7. The convergence of SPCA for Hippocampus image data.

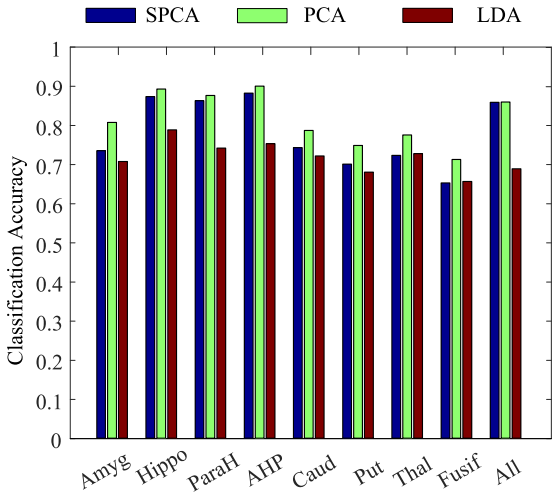


Fig. 8. The classification accuracy of three feature extraction methods with EN-ROC classifier.

Table 6
Feature dimensions after dimensionality reduction using different feature extraction methods.

	SPCA	PCA	LDA		SPCA	PCA	LDA
Amgy	13	310	71	Caud	25	310	60
Hippo	27	310	76	Put	18	310	70
ParaH	23	310	81	Thal	15	310	66
AHP	32	310	80	Fusif	31	310	92
All	87	310	116				

the image dimension is 17545 ($= 121 \times 145$). The learning curves is shown in Fig. 7, where 50 nonzero variables were selected from 17,545 variables randomly. As seen from Fig. 7, the SPCA is converged within 40 iterative steps. Then, without loss of generality, the iterative step of SPCA was set as 100 in our experiments.

Finally, we compared the performance of feature extraction between SPCA and other feature extraction methods. Compared with SPCA, PCA and LDA are all linear feature dimensionality reduction methods, which are popular feature extraction methods in machine learning. The feature dimensions of original sample data is 17545. After feature dimensionality reduction using SPCA, PCA and LDA respectively, Table 6 shows that SPCA and LDA only reserve 0.5% or less of the original feature dimensions while PCA still retains the sample size ($n = 310$) as feature dimensions. Based on the comparison results between four classifiers in Tables 3–5, we used EN-ROC as classifier after feature extraction procedure. As Fig. 8 shows, features extracted by PCA and SPCA perform about 10% better than those extracted by LDA. The identification accuracy of PCA or SPCA is over 15% higher than LDA in AHP and All GM ROIs. From Fig. 8, we also find that PCA performs 3.4% better than SPCA on average, which may be related to the information loss in sparsity. Especially, for some small ROIs (such as Amygdala), their local features can be protected well using compartmental feature dimension estimation method as shown in Fig. 6 while some information may be lost due to sparsity. This point is illustrated in Fig. 9, which is the average result of 50 times Monte-Carlo simulations. During each Monte-Carlo simulation, we used the volumes of 47 ROIs as variables for PCA and SPCA methods. And we hope some distinct ROIs can be derived as important variables for AD identification using feature extraction methods. In Fig. 9, the first five PCs using PCA method together explain 96.9% of the total variance, i.e. PCA reserves 96.9% energy of original data, while SPCA only reserves

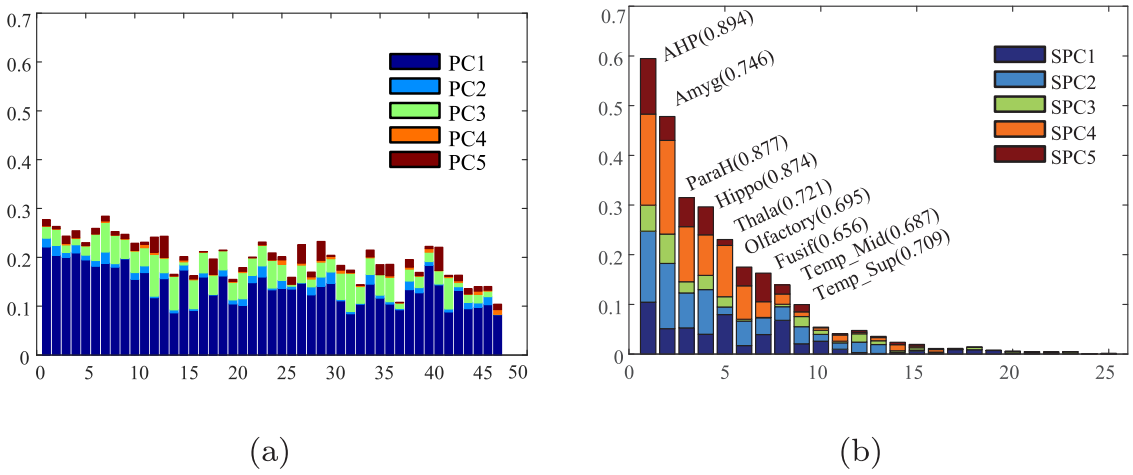


Fig. 9. The first five PCs and SPCs by PCA and SPCA respectively. (a) PCA. (b) SPCA. The vertical axes of (a) and (b) represent PCs values and SPCs values respectively. The number 1 ~ 47 in the horizontal axis in (a) represent 47 ROIs, which include 45 derived AAL ROIs and 2 combined ROIs (i.e. AHP and whole cerebral GM). The number 1 ~ 25 in the horizontal axis in (b) represent 25 selected ROIs using SPCA method. The first selected 9 ROIs names and the corresponding volume-based classification accuracy are listed in (b).

that of 32.9% with the first five SPCs. However, as what has been explained in Section 2, it is difficult for PCA to interpret the derived PCs while SPCA has the ability of group-variables selection. We can see from Fig. 9 that SPCA selected some important ROIs for AD identification while PCA cannot do that. For SPCA method, 25 ROIs were selected for AD identification according to analyzing the extracted SPCs. It is noticeable that some distinct ROIs (such as Amygdala, Hippocampus, Parahippocampal gyrus and AHP) were selected as the most important variables for AD identification, which is consistent with the clinical experience. More specifically, these distinct ROIs are correlated with each other, i.e. the variables for AD identification are correlated. This indicates that SPCA can identify important variables even for correlated variables.

5. Conclusion

We have presented a compartmental sparse feature extraction and classification method for AD identification in this study. The proposed method has two novel contributions: 1) reasonable feature dimension estimation with high computational efficiency by using the introduced union-of-subspace representation model, and 2) compartmental feature extraction and classification by using SPCA combined with EN-ROC classifier. The proposed method improves the automation of feature extraction and the performance of classification while protecting local features for the specified ROIs of the brain with high computational efficiency. Experimental results show that the extracted feature parameters well describe the atrophy of ROIs of the brain as clinical parameters but show better performance in AD identification than clinical parameters. That could be useful for doctors finding new feature parameters besides clinical parameters to describe the changes of cerebral pathology and cognitive function in AD patients. In this work, 2-D compartmental feature extraction may lose some information for some small ROIs due to sparsity (as discussed in Section 4), which may be a limitation for identifying some small ROIs. In the future work, 3-D feature extraction work will be discussed to solve this problem.

Conflict of Interest Statement

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Region-of-Interest Based Sparse Feature Learning Method for Alzheimers Disease Identification”.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2019.105290

References

[1] G. Mckhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of Alzheimer's disease: report of the nincds-adrda work group under the auspices of department of health and human services task force on Alzheimer's disease, Neurology 34 (7) (1984) 939–944.
[2] E.D. Roberson, M. Lennart, 100 Years and counting: prospects for defeating Alzheimer's disease, Science 314 (5800) (2006) 781–784.
[3] I. Alvarez, J.M. Gorriz, J. Ramirez, M. Lopez, C.G. Puntonet, F. Segovia, Alzheimer's diagnosis using eigenbrains and support vector machines, Electron Lett 45 (7) (2009) 342–343.
[4] X. Zhu, H.I. Suk, S.W. Lee, D. Shen, Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification, IEEE Trans. Biomed. Eng. 63 (3) (2016) 607–618.
[5] H.I. Suk, S.W. Lee, D. Shen, Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis, Brain Structure and Function 221 (5) (2016) 2569–2587.
[6] J. Young, G. Ridgway, K. Leung, S. Ourselin, Classification of Alzheimer's disease patients with hippocampal shape, wrapper based feature selection and support vector machine, in: SPIE Medical Imaging, 8314, 2012, pp. 1–7.
[7] X. Liu, D. Tosun, M.W. Weiner, N. Schuff, Locally linear embedding (LLE) for MRI based Alzheimer's disease classification, Neuroimage 83 (2013) 148–157.
[8] C. Chu, A.L. Hsu, K.H. Chou, P. Bandettini, C. Lin, Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images, Neuroimage 60 (1) (2012) 59–70.
[9] E.E. Bron, M. Smits, W.J. Niessen, S. Klein, Feature selection based on the SVM weight vector for classification of dementia, IEEE J Biomed Health Inform 19 (5) (2015) 1617–1626.
[10] M. Liu, D. Zhang, E. Adeli, D. Shen, Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 63 (7) (2016) 1473–1482.
[11] J. Torrents-Barrena, P. Lazar, R. Jayapathy, M.R. Rathnam, B. Mohandhas, D. Puig, Complex wavelet algorithm for computer-aided diagnosis of Alzheimer's disease, Electron Lett 51 (20) (2015) 1566–1568.
[12] B. Jie, M. Liu, J. Liu, D. Zhang, D. Shen, Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease, IEEE Trans. Biomed. Eng. 64 (1) (2017) 238–249.
[13] J. Escudero, E. Ifeachor, J.P. Zajicek, C. Green, J. Shearer, Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease, IEEE Trans. Biomed. Eng. 60 (1) (2013) 164–168.
[14] S. Minhas, A. Khanum, F. Riaz, A. Alvi, S.A. Khan, A nonparametric approach for mild cognitive impairment to AD conversion prediction: results on longitudinal data, IEEE J Biomed Health Inform 21 (5) (2017) 1403–1410.

[15] R. Cuingnet, J.A. Glaunes, M. Chupin, H. Benali, O. Colliot, Spatial and anatomical regularization of SVM: ageneral framework for neuroimaging data, IEEE Trans Pattern Anal Mach Intell 35 (3) (2013) 682–696.
[16] R. Guerrero, C. Ledig, D. Rueckert, Manifold alignment and transfer learning for classification of Alzheimer's disease, in: International Workshop on Machine Learning in Medical Imaging, 2014, pp. 77–84.
[17] N.K. Batmanghelich, B. Taskar, C. Davatzikos, Generative-discriminative basis learning for medical imaging, IEEE Trans Med Imaging 31 (1) (2012) 51–69.
[18] F. Liu, L. Zhou, C. Shen, J. Yin, Multiple kernel learning in the primal for multimodal Alzheimer's disease classification, IEEE J Biomed Health Inform 18 (3) (2014) 984–990.
[19] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M.J. Fulham, Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1132–1140.
[20] S. Lahmiri, A. Shmuel, Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease, Biomed Signal Process Control 52 (2019) 414–419.
[21] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Adv Neural Inf Process Syst 18 (2005) 507–514.
[22] W.D. Penny, K.J. Friston, J.T. Ashburner, stefan Kiebel, T. Nichols, Statistical parametric mapping : The analysis of functional brain images, Elsevier, 2006.
[23] J. Wan, Z. Zhang, B.D. Rao, S. Fang, J. Yan, A.J. Saykin, L. Shen, Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation- and nonlinearity-aware sparse bayesian learning, IEEE Trans Med Imaging 33 (7) (2014) 1475–1487.
[24] S. Huang, J. Li, J. Ye, A. Fleisher, K. Chen, T. Wu, E. Reiman, A sparse structure learning algorithm for gaussian bayesian network identification from high-dimensional data, IEEE Trans Pattern Anal Mach Intell 35 (6) (2013) 1328–1342.
[25] S. Lahmiri, Image characterization by fractal descriptors in variational mode decomposition domain: application to brain magnetic resonance, Physica A Statistical Mechanics and Its Applications 456 (2016) 235–243.
[26] D. Pachauri, C. Hinrichs, M.K. Chung, S.C. Johnson, V. Singh, Topology-based kernels with application to inference problems in Alzheimer's disease, IEEE Trans Med Imaging 30 (10) (2011) 1760–1770.
[27] X. Tang, J. Wu, Principal component analysis of the shape deformations of the hippocampus in Alzheimer's disease, in: Engineering in Medicine and Biology Society, 2016, pp. 4013–4016.
[28] Y. Liu, C. Ma, B.A. Clifford, F. Lam, C.L. Johnson, Z.P. Liang, Improved low-rank filtering of magnetic resonance spectroscopic imaging data corrupted by noise and B_0 field inhomogeneity, IEEE Trans. Biomed. Eng. 63 (4) (2016) 841–849.
[29] L. Wang, Y. Liu, H. Cheng, X. Zeng, Z. Wang, Elastic net based sparse feature learning and classification for Alzheimer's disease identification, in: Engineering in Medicine and Biology Society, 2017, pp. 2288–2291.
[30] Y. Liu, L. Wang, X. Zeng, Z. Wang, Y. Gao, Compartmental sparse feature selection method for Alzheimer's disease identification, in: Engineering in Medicine and Biology Society, 2017, pp. 3073–3076.
[31] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, Journal of Computational and Graphical Statistics 15 (2) (2006) 265–286.
[32] F. Li, L. Tran, K.H. Thung, S. Ji, D. Shen, J. Li, Robust deep learning for improved classification of AD/MCI patients, in: 5th International Workshop, MLMI, 2014, pp. 240–247.
[33] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: IEEE International Symposium on Biomedical Imaging, 2014, pp. 1015–1018.
[34] J.M. Mateos-Perez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, A.C. Evans, Structural neuroimaging as clinical predictor: a review of machine learning applications, NeuroImage: Clinical (2018) 506–522.
[35] K.J. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline, C.D. Frith, R.S.J. Frackowiak, Statistical parametric maps in functional imaging: ageneral linear approach, Hum Brain Mapp 2 (4) (1994) 189–210.
[36] J.A. Maldjian, P.J. Laurienti, R.A. Kraft, J.H. Burdette, An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets, Neuroimage 19 (3) (2003) 12–33.
[37] N. Tzouriomazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, Neuroimage 15 (1) (2002) 273–289.
[38] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.
[39] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society 67 (2) (2005) 301–320.
[40] P. Padilla, M. Lopez, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, I. Alvarez, NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease, IEEE Trans Med Imaging 31 (2) (2012) 207–216.
[41] J.H. Morra, Z. Tu, L.G. Apostolova, A.E. Green, A.W. Toga, P.M. Thompson, Comparison of adaboost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation, IEEE Trans Med Imaging 29 (1) (2010) 30–43.
[42] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans Pattern Anal Mach Intell 31 (2) (2009) 210–227.
[43] Z. Wang, P.J. Markiewicz, G. Platsch, J. Kornhuber, T. Kuwert, D. Merhof, Comparison of methods for classification of Alzheimer's disease, frontotemporal dementia and asymptomatic controls, in: Nuclear Science Symposium and Medical Imaging Conference, 2013, pp. 1–6.
[44] Y. Liu, X. Zeng, L. Wang, Z. Wang, Q. Wang, Comparison and analyzation of different feature parameters for Alzheimer's disease identification, in: International Conference of the IEEE Engineering in Medicine and Biology Society, 2019, pp. 1–4.
[45] A.J. Bastos Leite, P. Scheltens, F. Barkhof, Pathological aging of the brain: an overview, Topics in Magnetic Resonance Imaging 15 (6) (2004) 369–389.
[46] G. Chetelat, B. Desgranges, L.S.V. De, F. Viader, F. Eustache, J.C. Baron, Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment, Neuroreport 13 (15) (2002) 1939–1943.
[47] C. Pennanen, C. Testa, M.P. Laakso, M. Hallikainen, E.L. Helkala, T. Hänninen, M. Kivipelto, M. Könönen, A. Nissinen, S. Tervo, A voxel based morphometry study on mild cognitive impairment, Journal of Neurology Neurosurgery and Psychiatry 76 (1) (2005) 11–14.