Sparse Kernel Entropy Component Analysis for Dimensionality Reduction of Neuroimaging Data

Qikun Jiang, Jun Shi

Abstract—**The neuroimaging data typically has extremely high dimensions. Therefore, dimensionality reduction is commonly used to extract discriminative features. Kernel entropy component analysis (KECA) is a newly developed data transformation method, where the key idea is to preserve the most estimated Renyi entropy of the input space data set via a kernel-based estimator. Despite its good performance, KECA still suffers from the problem of low computational efficiency for large-scale data. In this paper, we proposed a sparse KECA (SKECA) algorithm with the recursive divide-and-conquer based solution, and then applied it to perform dimensionality reduction of neuroimaging data for classification of the Alzheimer's disease (AD). We compared the SKECA with KECA, principal component analysis (PCA), kernel PCA (KPCA) and sparse KPCA. The experimental results indicate that the proposed SKECA has most superior performance to all other algorithms when extracting discriminative features from neuroimaging data for AD classification.**

I. INTRODUCTION

The neuroimaging techniques, such as magnetic resonance imaging (MRI), functional MRI, and positron emission tomography, are commonly applied to human brain imaging. Furthermore, the neuroimaging data based computer-aided methods have been proved very helpful for diagnosing brain disease diagnosis, such as Alzheimer's disease (AD) [1] and Parkinson's disease [2].

The neuroimaging data is of extremely high dimension but with small sample size, which will degrade classification performance. Therefore, dimensionality reduction methods have been used to overcome this problem for neuroimaging data, such as principal component analysis (PCA) [3], kernel PCA (KPCA) [4], nonnegative matrix factorization [5], and Laplacian Eigenmaps [6].

PCA is a well-known linear feature extraction method and widely used in different applications. KPCA is the nonlinear extension of PCA with kernel method, which also performs well. Both PCA and KPCA perform dimensionality reduction by selecting the top eigenvalues and their corresponding eigenvectors. However, the resulting transformation may be based on uninformative eigenvectors from the viewpoint of information theory [7].

The kernel entropy component analysis (KECA) is a newly developed data transformation method, whose its key idea is based on preserving the most estimated Renyi entropy of the input space data set via a kernel-based estimator [8]. Consequently, KECA doesn't necessarily select the top eigenvalues and eigenvectors of the kernel matrix, but still reveals intrinsic structure related to information entropy of the input space data set [8]. Moreover, KECA typically produces a transformed data set with a distinct angular structure, which benefits the further signal processing [8]. Therefore, KECA has been successfully applied for data clustering [8][9], feature extraction [7][10], etc.

Despite its many advantages, KECA still suffers from one important disadvantage that it might be quite inefficient when processing large-scale samples. Similar to KPCA, KECA requires evaluation of the kernel function in respect of all training samples when computing PC projection for a given input. This is unfortunate limitations, because the larger the size of the training samples, the lower the computational efficiency of KECA.

Accordingly, a natural approach to improve the computational efficiency of PCA or KPCA is to integrate sparse solution in it, which has attracted much attention. Consequently, many sparse PCA (SPCA) and sparse KPCA (SKPCA) algorithms have been proposed in recent years [11][12][13][14][15][16][17][18]. However, to the best of our knowledge, sparse KECA (SKECA) algorithm has not been proposed yet, which in fact is very necessary for large-scale neuroimage data processing.

There are mainly two methodologies to realize SPCA [17]. One is the greedy approach focusing on the solving of one-sparse-PC model, and the other is the block approach aiming to calculate multiple sparse PCs at once by utilizing certain block optimization techniques [17]. More recently, a recursive divide-and-conquer (ReDaC) based method was proposed to solve the SPCA problem [17]. It decomposes the original large and complex problem of PCA into a series of small and simple sub-problems, and then recursively solve them. The experimental results indicate the effectiveness of this SPCA algorithm.

In this paper, we proposed a novel SKECA algorithm motivated by the ReDac solution [17], and then applied it to reduce the feature dimensions from neuroimaging data to discriminate AD from normal subjects.

This research is supported by the Shanghai Municipal Natural Science Foundation (12ZR1410800) and the Innovation Program of Shanghai Municipal Education Commission (13YZ016).

Qikun Jiang is with the School of Communication and Information Engineering, Shanghai University, Shanghai, China (e-mail: jiangqikun@shu.edu.cn).

Jun Shi is with the School of Communication and Information Engineering, Shanghai University, Shanghai, China (+86-21-66137256; e-mail: junshi@staff.shu.edu.cn).

II. METHOD

A. Kernel entropy component analysis

Kernel entropy component analysis is very closely related to the KPCA, but it focuses on entropy components instead of principal components in PCA and KPCA.

The derivation of KECA is not very similar to KPCA. It starts by expressing an estimate of the continuous Renyi (quadratic) entropy based on kernel density estimator, but it is actually quite surprising that it leads to a method so similar to KPCA, The briefly introduction of KECA is as following [8].

Let: $R^d \rightarrow F$ denote a nonlinear map such that $x_t \rightarrow \varphi(x_t)$, and let $\boldsymbol{\varphi}(x)=[\varphi(x_1),\ldots,\varphi(x_N)]$. Inner- products in the Hilbert space F can be computed via a positive semidefinite Mercer's kernel function K: $R^d \times R^d \rightarrow R$

$$\mathrm{K}(x_t, x_{t*}) = \langle \varphi(x_t), \varphi(x_{t*}) \rangle \tag{1}$$

Defining ($N\times N$) the Mercer kernel matrix $\mathbf{K}$ such that element (t,t^*) of $\mathbf{K}$ equals $k(x_t,x_t^*)$, then $\mathbf{K}=\boldsymbol{\varphi}^{\mathbf{T}}\boldsymbol{\varphi}$ is an inner-product (Gram) matrix in F. The kernel matrix can be eigendecomposed as $\mathbf{K}=\mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}}$, where $\mathbf{D}$ is a diagonal matrix storing the eigenvalues $\lambda_1,\ldots,\lambda_N$ and $\mathbf{E}$ is a matrix with the corresponding eigenvectors $e_1,\ldots,e_N$ as columns.

A projection of $\boldsymbol{\varphi}$ onto a single principal axis is given by $u_i^{\mathrm{T}}=\lambda_i^{1/2}e_i^{\mathrm{T}}$. Hence these projections of $\boldsymbol{\varphi}$ onto all principal axes are given by $\mathbf{U}^{\mathrm{T}}\boldsymbol{\varphi}=\mathbf{D}^{1/2}\mathbf{E}^{\mathrm{T}}$, where $\mathbf{U}=[u_1,\ldots,u_N]$ is the projection matrix.

We define KECA as a k-dimensional data transformation obtained by projecting $\boldsymbol{\varphi}$ onto a subspace $\mathbf{U}_k$ spanned by those k feature space principal axes contributing most to the Renyi entropy estimate of data, obtaining the extracted KECA features

$$\boldsymbol{\varphi}_{eca} = \mathbf{U}_k^{\mathrm{T}}\boldsymbol{\varphi} = \mathbf{D}_k^{1/2}\mathbf{E}_k^{\mathrm{T}} \tag{2}$$

this is the solution to the minimization problem

$$\begin{aligned}\boldsymbol{\varphi}_{eca} &= \mathbf{D}_k^{1/2}\mathbf{E}_k^{\mathrm{T}} \min_{\lambda_1,e_1,\ldots,\lambda_N,e_N} \hat{\mathrm{V}}(\mathrm{p}) - \hat{\mathrm{V}}_k(\mathrm{p}) \\ &= \frac{1}{N^2}\sum_{i=k+1}^{N}\psi_i\end{aligned} \tag{3}$$

where the entropy estimate associated with $\boldsymbol{\varphi}_{\mathrm{eca}}$ is

$$\hat{\mathrm{V}}_k(\mathrm{p}) = \frac{1}{N^2}\sum_{i=1}^{k}(\sqrt{\lambda_i}\,e_i^{\mathrm{T}}1)^2 = \frac{1}{N^2}\sum_{i=1}^{k}\psi_i \tag{4}$$

each term ψ_i in this expression will contribute to the entropy estimate. We select eigenvalues and corresponding eigenvectors which contribute more to the entropy estimate. Note that the KPCA transformation is based solely on the top eigenvalues of $\mathbf{K}$ and will, in general, differ from KECA.

For out-of-sample data point $\varphi(x)$, we can obtain mapped feature

$$\boldsymbol{\varphi}_{eca}^{\ *} = \mathbf{U}_k^{\mathrm{T}}\boldsymbol{\varphi}^* = \mathbf{D}_k^{-1/2}\mathbf{E}_k^{\mathrm{T}}\mathbf{K}^* \tag{5}$$

where $\mathbf{K}^*=\boldsymbol{\varphi}^{\mathbf{T}}\boldsymbol{\varphi}^*$ and $\boldsymbol{\varphi}^*$ refer to a collection of out-of-sample data points. Furthermore, (2) can be written as:

$$\boldsymbol{\varphi}_{eca} = \mathbf{U}_k^{\mathrm{T}}\boldsymbol{\varphi} = \mathbf{D}_k^{1/2}\mathbf{E}_k^{\mathrm{T}} = \mathbf{D}_k^{-1/2}\mathbf{E}_k^{\mathrm{T}}\mathbf{K} \tag{6}$$

B. Sparse Kernel Entropy Component Analysis

The SPCA problem has the following two mathematical formulations [17]:

$$\max_{\mathbf{V}} \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\Sigma\mathbf{V})\ s.t.\ v_i^{\mathrm{T}}v_i = 1, \|v_i\|_p \le t_i\,(i=1,\ldots,k) \tag{7}$$

and

$$\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}\right\|_F^2\ s.t.\ v_i^{\mathrm{T}}v_i = 1, \|v_i\|_p \le t_i\,(i=1,\ldots, k) \tag{8}$$

where $\mathbf{X}=[x_1,\ldots,x_N]^{\mathrm{T}}\in R^{N\times d}$, $\mathbf{V}\in R^{d\times k}$ is the matrix of principal component loading array. $\mathbf{U}\in R^{N\times k}$ is the matrix of projected data.

The ReDaC method which depend on the second model(8) can easily separate the original large problem into a series of small problems, has proved its efficiency in SPCA [17]. Therefore, we employ this method to our proposed SKECA. Rewriting (3) as follows:

$$\begin{aligned}\boldsymbol{\varphi}_{eca} &= \mathbf{D}_k^{1/2}\mathbf{E}_k^{\mathrm{T}} \min_{\lambda_1,e_1,\ldots,\lambda_N,e_N} \hat{\mathrm{V}}(\mathrm{p}) - \hat{\mathrm{V}}_k(\mathrm{p}) \\ &= \min_{\lambda_1,e_1,\ldots,\lambda_N,e_N} \frac{1}{N^2}\mathbf{1}^{\mathrm{T}}(\mathbf{K}-\mathbf{K}_{eca})\mathbf{1}\end{aligned} \tag{9}$$

where $\mathbf{K}_{eca}=\mathbf{E}_k\mathbf{D}_k\mathbf{E}_k^{\mathrm{T}}$. The purpose of SKECA is obtaining sparse projection matrix: $\mathbf{E}_k$ of (2) is sparse. In this paper, we employ the ReDaC method to achieve this goal. We give the solution to the minimization problem of SKECA as follows:

$$\begin{aligned}\min_{\mathbf{K}_{eca}} \left\|\mathbf{K}-\mathbf{K}_{eca}\right\|_F^2 &= \min_{\mathbf{E}_k,\mathbf{D}_k} \left\|\mathbf{K}-\mathbf{E}_k\mathbf{D}_k\mathbf{E}_k^{\mathrm{T}}\right\|_F^2 \\ &= \min_{\mathbf{E}_k,\mathbf{D}_k} \left\|\mathbf{K}-\mathbf{U}\mathbf{V}^{\mathrm{T}}\right\|_F^2 \\ s.t.\ v_i^{\mathrm{T}}v_i &= 1, \|v_i\|_p \le t_i\,(i=1,\ldots, k)\end{aligned} \tag{10}$$

where $\mathbf{U}=\mathbf{E}_k\mathbf{D}_k$ and $\mathbf{V}=\mathbf{E}_k$. Therefore, the objective function of the (10)can be formulated as follows:

$$\begin{aligned}\left\|\mathbf{K}-\mathbf{U}\mathbf{V}^{\mathrm{T}}\right\|_F^2 &= \left\|\mathbf{K}-\sum\nolimits_{j=1}^{k} u_j v_j^{\mathrm{T}}\right\|_F^2 \\ &= \left\|\mathbf{E}_i - u_i v_i^{\mathrm{T}}\right\|_F^2\end{aligned} \tag{11}$$

where $\mathbf{E}_i = \mathbf{K} - \sum\nolimits_{j\ne i} u_j v_j^{\mathrm{T}}$. It is then easy to separate the original large minimization problem into a series of small minimization problems, which are each with respect to a column vector u_i of $\mathbf{U}$ and v_i of $\mathbf{V}$ for $i=1,\ldots,k$, respectively, as follows:

$$\min_{v_i} \left\| \mathbf{E}_i - u_i v_i^{\mathrm{T}} \right\|_F^2$$
$$s.t.\ v_i^{\mathrm{T}} v_i = 1, \left\| v_i \right\|_p \le t_i \ (i = 1, \ldots, k) \quad (12)$$

and

$$\min_{u_i} \left\| \mathbf{E}_i - u_i v_i^{\mathrm{T}} \right\|_F^2 \quad (13)$$

Through recursively optimizing these small sub-problems, the ReDac method for solving the minimization problem of SKECA can then be naturally constructed.

The calculation of **U** and **V** is to recursively optimize each column, u_i of **U** and v_i of **V** for $i=1,\ldots,k,$ with other u_j and v_j fixed. By iteratively implementing the above procedures, **U** and **V** can be recursively updated until the stopping criterion is satisfied.

We summarize SKECA algorithm as follows: KECA first calculates the aforementioned kernel matrix **K** using the given input data set $\mathbf{X} \in R^{d\times N}$, and then gets **U** and **V** by $\mathbf{U}=\mathbf{E}_k\mathbf{D}_k$ and $\mathbf{V}= \mathbf{E}_k$; finally, obtains sparse projecting matrix **V** through ReDaC method. Thus, the transformed result of SKECA can be achieved through following formula:

$$\boldsymbol{\varphi}_{skeca} = \mathbf{U}_k^{\mathrm{T}} \boldsymbol{\varphi} = \mathbf{D}_k^{1/2} \mathbf{V}^{\mathrm{T}} = \mathbf{D}_k^{-1/2} \mathbf{V}^{\mathrm{T}} \mathbf{K} \quad (14)$$

where $\mathbf{U}_k =[u_1,\ldots,u_k]$, $u_i^{\mathrm{T}}=\lambda_i^{1/2} v_i^{\mathrm{T}}$.

Based on the ReDaC method, we also proposed a SKPCA algorithm with similar solution for comparison study. The main difference between SKECA and SKPCA is the initialization of **U** and **V**. Although the calculation formulations of initializations of **U** and **V** are same, which are $\mathbf{U}=\mathbf{E}_k\mathbf{D}_k$ and $\mathbf{V}=\mathbf{E}_k$, $\mathbf{D}_k$ and $\mathbf{E}_k$ are different because they depend on the selection of different eigenvalues and eigenvectors in SKECA or SKPCA.

III. Experiment

A. Data

To evaluate the performance of proposed SKECA algorithm, we applied it to neuroimaging data for dimensionality reduction. We selected the MR images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [19]. In this paper, only ADNI subjects with all corresponding MRI, CSF and PET baseline data were included, which has been used in reference [20]. This yields a total of 103 subjects including 51 AD patients and 52 healthy controls (HC). For simplification, we only selected MR data for testing. The data pre-processing and feature extraction were same as in reference [20]. Specifically, we did anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). With atlas warping, each MR image was partitioned into 93 region of interests (ROIs), and for each of the 93 ROIs, the volumes of gray matter tissue were calculated as a feature. Consequently, totally 93 features were extracted from each MR image.

B. Experimental Setup

We compared the proposed SKECA with the proposed SKPCA ,SPCA [17], original KECA, KPCA and PCA. It is worth noting that the 4-order polynomial kernel was used for all kernel-based methods. To evaluate the performances of different algorithms, the extracted features by different methods were fed to k-nearest neighbor (KNN) classifier to discriminate AD from HC. As the reduced features by KECA and SKECA has a distinct angular structure, the Cosine similarity was applied in KNN for KECA and SKECA, while the Euclidean distance was used in KNN for other algorithms. The KNN classifier with different distance measure is to secure the best performances of different methods.

The 10-fold cross-validation strategy was performed on 51 AD patients and 52 HC subjects. The classification accuracy and sensitivity were selected as evaluation indices. A paired-samples *t*-test was used to statistically evaluate the performances between the proposed SKECA and other dimensionality reduction algorithms. The results were declared statistically significant when associated with *p*-value that is less than 0.05.

IV. Experimental Results

Table 1 and Tables 2 show the classification results of different feature extraction algorithms at different feature dimensions. It can be found that the best classification accuracy and sensitivity are 91.47±0.82% and 92.00±2.67% by SKECA with 35 features. The SPCA ranks second only to SKECA with the classification accuracy of 90.16±0.87% and sensitivity of 86.03±2.33%. SKECA achieved at least 1% and 4% improvements on accuracy and sensitivity. We also calculated the *t*-test between the best results of SKECA with those of other algorithms, and the proposed SKECA algorithm significantly improves the performance of discriminating AD from HC subjects, compared to other algorithms with all the *p*-value less than 0.05. It is also worth noting that all the sparsity-based algorithms significantly outperform the original algorithms without sparsity embedded, which indicates that sparse representation really improve the performance.

In this work, the 93 ROIs were used as the original features. However, they are not equally important to represent AD. Therefore, it is necessary to further reduce feature dimensionality. The results indicate that SKECA works well, which is mainly due to the following two factors: the distinct angular structure of transformed data set and the denoising ability introduced by sparsity. The angular structure of reduced features makes the classification more easily with Cosine similarity in KNN. Moreover, the noise is still inevitable, though the MRI data in our experiment have been pre-processed. Hence, the dimensionality reduction algorithms used for neuroimaging data should have strong robust ability for noise. The sparsity in SKECA not only improves the computational efficiency, but also makes SKECA more robust against noise. The reason is that the output transformed data of SKECA can be regarded as the weighted sum of kernel matrix **K** with sparse coefficient matrix **V**. Therefore, the more relative data are selected for calculation of inner product, which reduces the noise.

TABLE I. CLASSIFICATION ACCURACIES OF DIFFERENT DIMENSIONALITY REDUCTION ALGORITHMS (UNIT: %)

Algorithm \ Dimension	10	15	20	25	30	35	40
PCA	**75.12±2.45**	73.32±3.30	72.73±2.88	72.35±2.60	72.55±2.26	72.64±2.75	73.75±2.56
KPCA	**70.73±3.49**	65.99±2.36	67.53±2.36	69.98±3.14	69.80±2.26	66.95±3.47	66.66±2.98
KECA	72.29±1.84	71.41±2.75	72.13±2.87	**73.01±3.23**	72.42±1.81	71.22±2.68	70.74±3.75
SPCA	**90.16±0.87**	88.96±0.79	88.69±1.08	88.27±0.73	87.33±0.78	87.46±1.18	86.00±0.72
SKPCA	82.94±1.59	81.05±1.53	82.90±1.86	83.63±1.58	**85.43±1.53**	84.70±1.40	83.27±1.60
SKECA	86.96±0.88	88.19±1.51	88.31±1.24	88.68±1.09	89.38±0.75	**91.47±0.82**	90.14±0.90

TABLE II. CLASSIFICATION SENSITIVITIES OF DIFFERENT DIMENSIONALITY REDUCTION ALGORITHMS (UNIT: %)

Algorithm \ Dimension	10	15	20	25	30	35	40
PCA	66.43±3.02	68.23±4.03	67.50±2.99	**69.00±4.02**	66.83±3.70	68.57±4.88	66.70±3.82
KPCA	**70.00±4.22**	62.07±3.23	59.20±4.06	62.03±4.59	63.27±3.36	57.97±4.80	55.13±4.75
KECA	69.63±4.17	73.33±4.10	75.17±3.88	**75.87±4.53**	75.73±3.69	73.20±2.99	75.43±2.60
SPCA	**86.03±2.33**	84.20±2.68	84.03±1.96	84.03±1.71	82.50±3.78	83.33±3.56	81.20±2.03
SKPCA	**81.93±3.87**	77.40±3.71	77.43±2.74	78.50±2.67	81.70±4.03	77.67±3.77	76.27±3.48
SKECA	86.00±2.42	89.00±2.62	88.73±2.91	89.17±2.65	89.40±2.31	**92.00±2.67**	91.07±2.30

V. CONCLUSION

In this paper, we proposed a sparse KECA algorithm for dimensionality reduction, and then applied it to MR images to extract discriminative features for classification of AD. The experimental results indicate that our proposed algorithm can significantly improve the performance of dimensionality reduction, leading to better classification performance compared with other algorithms. In the future work, more high dimensional neuroimaging data will be tested by the proposed SKECA algorithm. We will also further improve the proposed SKECA to have high classification performance with less feature dimensions.

REFERENCES

[1] G. Orru, W. Pettersson-Yeo, A. F. Marquand, et al. "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review," *Neurosci. Biobehav. R.*, vol. 36, pp. 1140-1152, 2012.

[2] A. F. Marquand, M. Filippone, J. Ashburner, et al. "Automated, high accuracy classification of Parkinsonian disorders: a pattern recognition approach," *PLoS One*, 2013.

[3] M. Lopez, J. Ramirez, J. M. Gorriz, et al. "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," *Neurocomputing*, vol. 74, pp. 1260-1271, 2011.

[4] M.M. Lopez, J. Ramirez, J.M. Gorriz, et al. "SVM-based CAD system for early detection of the alzheimer's disease using kernel PCA and LDA," *Neuro. Sci. Lett.*, vol. 464, pp. 233-238, 2009.

[5] P. Padilla, M. Lopez, J. M. Górriz, et al. "NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease," *IEEE Trans. Med. Imaging*, vol. 31, pp. 207-216, 2012.

[6] R. Wolz, P. Aljabar, J. V. Hajnal, et al. "Nonlinear dimensionality reduction combining MR imaging with non-imaging information," *Med. Image. Anal.*, vol. 16, pp. 819-830, 2012.

[7] Z. H. Zhang, and E. R. Hancock, "Kernel entropy-based unsuprevised spectral feature selection," *Int. J. Pattern Recogn. Artif. Intell.*, vol. 26, 2012.

[8] R. Jenssen, "Kernel entropy component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 847-860, 2010.

[9] L. Gómez-Chova, R. Jenssen, and G. Camps-Valls. "Kernel entropy component analysis for remote sensing image clustering," *IEEE Geosci. Remo. S. Lett.*, vol. 9, pp. 312-316, 2012.

[10] B. H. Shekar, M. S. Kumari, L. M. Mestetskiy, et al. "Face recognition using kernel entropy component analysis," *Neurocomputing*, vol. 74, pp. 1053-1057, 2011.

[11] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Coput. Graph. Stat.,* vol. 15, pp. 265-286, 2006.

[12] F. R. Bach , L. E. Ghaoui and A. d'Aspremont, "Full regularization path for sparse principal component analysis," In *Proc. 24th Int. Conf. Mach. Learn,* pp. 177-184,2007.

[13] H. P. Shen, and J. h. Z. Huang. "Sparse principal component analysis via regularized low rank matrix approximation," *J. Multivariate. Anal.*, vol. 99, pp. 1015-1034, 2008.

[14] M. Journee, Y. Nesterov, P. Richtarik, and R. Sepulchre, "Generalized power method for sparse principal component analysis," *J. Mach. Learn. Res.*, vol. 11 pp. 517-553, 2010.

[15] Y. W. Zhang and L. E. Ghaoui. "Large-scale sparse principal component analysis with application to text data," In *Proc. Adv. NIPS.*, Dec 2011.

[16] Z. M. Ma. "Sparse principal component analysis and iterative thresholding," *Ann. Stat.*,Vol. 41, No. 2, pp.772-801,2013

[17] Q. Zhao, D. Y. Meng, and Z. B. Xu. "A recursive divide-and-conquer approach for sparse principal component analysis," *arXiv*: 1211.7219, 2013.

[18] M. E. Tipping. "Sparse kernel principal component analysis," *J. Mach. Learn. Res.*, vol. 1, pp. 211-244, 2001.

[19] C. R. Jack, M. A. Bernstein, N. Fox, et.al. "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imaging*, vol. 27, pp. 685-691, 2008.

[20] D. Q. Zhang, Y. P. Wang, and L. P. Zhou, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, pp. 856-867, 2011.