

A SUPERVISED MULTIVIEW SPECTRAL EMBEDDING METHOD FOR NEUROIMAGING CLASSIFICATION

Sidong Liu¹, Lelin Zhang¹, Weidong Cai¹, Yang Song¹, Zhiyong Wang¹,
Lingfeng Wen², David Dagan Feng¹

¹Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia

²Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

ABSTRACT

The multi-view/multi-modal features are commonly used in neuroimaging classification because they could provide complementary information to each other and thus result in better classification performance than single-view features. However, it is very challenging to effectively integrate such rich features, since straightforward concatenation or single-view spectral embedding methods rarely leads to physically meaningful integration. In this paper, we present a supervised multi-view/multi-modal spectral embedding method (SMSE) for neuroimaging classification. This method embeds the high dimensional multi-view features derived from multi-modal neuroimaging data into a low dimensional feature space and preserves the optimal local embeddings among different views. The proposed SMSE algorithm, validated using three groups of neuroimaging data, is able to achieve significant classification improvement over the state-of-the-art multi-view spectral embedding methods.

Index Terms— supervised learning, multi-view spectral embedding, neuroimaging classification

1. INTRODUCTION

Neuroimaging, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), is a fundamental component in the diagnosis of neurological disorders and also an important indicator in disease monitoring and therapy assessments.

A variety of neuroimaging classification methods for the diagnosis of neurological disorders, such as Alzheimer's disease (AD) and mild cognitive impairment (MCI), have been proposed [1-12]. Single-view features, *e.g.*, texture features [1, 2], geometric features [3, 4], volumetric features [5, 6], and physiological features [7, 8] were usually used in these studies. There were also studies on the multi-view/multi-modal features [9-12]. Multi-view/modal features could depict neuroimaging data from different perspectives and therefore provide complementary information to each other. The primary features were

usually extracted from the MRI data [9-11] and/or PET data [10, 12], and sometimes combined with others non-imaging features, *e.g.*, cerebrospinal fluid (CSF) measures [9] and genetic biomarkers [11]. The naive concatenated features were subsequently used to train classifiers, *e.g.*, support vector machines (SVM). However, the concatenated features are not physically meaningful and such high dimensional features could lead to the "curse of dimensionality". In several studies [10-12], a feature selection step was thus applied to reduce the feature dimensionality based on the assumption that certain features hardly affected the classification performance and therefore could be discarded. However, we believe such assumptions were not convincing because it is difficult to compare the multi-view features on the same basis and the grouping effects of features, which are important in neuroimaging classification, are usually ignored in feature selection.

Other than concatenating the multi-modal/multi-view features into a high dimensional feature vector or selecting a subset of features to represent the subjects, we may embed the high dimensional features into a low dimensional space by spectral embedding. Such embeddings may reduce the complexity of high dimensional feature space without discarding less important features. However, the existing spectral embedding algorithms, such as principal component analysis (PCA), Laplacian eigenmaps (LE) [13] and ISOMAP [14], assume the data are distributed in a unified feature space; therefore they could not be effectively used in multi-view features embedding.

Recently Long *et al.* [15] and Xia *et al.* [16] attempted to solve the multi-view embedding problems using the distributed spectral embedding (DSE) and multi-view spectral embedding (MSE) algorithms, respectively. DSE assumes the low-dimensional embedding of each view is known and learns a consensus embedding based on the individual embeddings. MSE preserves the local neighborhood embeddings and seeks a global alignment wherein the embeddings of all views are sufficiently smooth. MSE also allows to set different weights to individual views. Both DSE and MSE focus on the unsupervised classification that the label/diagnosis of each subject is not used in finding the embeddings. However, we

believe such information could further facilitate neuroimaging classification. Therefore, in this study, we proposed a supervised multi-view spectral embedding (SMSE) method for neuroimaging classification. This method is based on MSE and incorporates the supervised learning in deriving the local neighborhood embeddings. We validated the SMSE algorithm using the multi-view features extracted from MRI and PET data of three groups of subjects, and the preliminary results showed that SMSE had a great potential to enhance neuroimaging classification.

2. METHODS

2.1 Image Acquisition and Pre-processing

The neuroimaging data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiatives (ADNI) database (adni.loni.ucla.edu) [17]. In total, 331 subjects were randomly selected from the ADNI baseline cohort, and a T1-weighted 1.5-Tesla MRI image and an FDG-PET image were retrieved for each subject. The sample dataset included three groups of subjects, including 85 AD cases, 169 MCI cases and 77 cognitive normal (CN) subjects.

All 3D MRI and PET data were converted to the ADNI format following the ADNI image correction protocol [17, 18]. The PET images were further aligned to the corresponding MRI image using FSL FLIRT [19]. We then nonlinearly registered the MRI images to the well-known ICBM_152 template [20] consisting of 83 brain functional regions using the Image Registration Toolkit (IRTK) [21]. The outputted registration coefficients by IRTK were applied to warp the aligned PET images into the template space. We finally mapped all brain functional regions in the template space using the multi-atlas propagation with enhanced registration (MAPER) approach [22] on each registered MRI and PET image for the region-wise feature extraction.

2.2 Multi-view Feature Extraction

Totally four views of features were extracted from the PET and MRI data, namely the cerebral metabolic rate of glucose consumption (CMRGlc), the Grey Matter (GM) volume fractions, solidity features, and convexity features.

CMRGlc is the most widely used feature for PET data in AD and MCI diagnosis [10, 12]. The CMRGlc parameters could reflect the metabolic rate of the brain regions and detect the early functional changes in brain. The average CMRGlc parameters were extracted from V ($V = 83$) brain regions from each registered PET image (indexed by j) as the first view of features, $X^{(1)} = \{x_1^{(1)}, \dots, x_j^{(1)}, \dots, x_N^{(1)}\}_{j=1}^N$, $X^{(1)} \in \mathbb{R}^{V \times N}$ ($N = 331$).

The other three features were extracted from MRI data to capture the structural variations. GM volume features

were extracted from the V brain regions for each subject as the second view of features $X^{(2)} = \{x_1^{(2)}, \dots, x_N^{(2)}\}$. We further normalized $x_n^{(2)}(v)$, where $v = 1, \dots, V$, for each brain region (indexed by v) as a fraction of the whole brain. Two other views, the convexity ($X^{(3)} \in \mathbb{R}^{V \times N}$) and solidity ($X^{(4)} \in \mathbb{R}^{V \times N}$) were also extracted in addition to the GM volume fractions. Both convexity and solidity required the convex hull. The $x_n^{(3)}(v)$ was defined as the ratio of the convex hull surface area to the surface area of v^{th} brain region; and the $x_n^{(4)}(v)$ was the ratio of the volume of v^{th} brain region to the volume of the convex hull. The convexity and solidity were used to describe the shape features of the brain regions.

2.3 Supervised Multi-View Spectral Embedding

The original MSE focuses on unsupervised embedding which does not incorporate the label/diagnosis information of subjects while deriving the local embeddings. Based on the MSE framework [16], we adopted the supervised learning in local neighborhood embedding. The global alignment based on the derived local embeddings was then calculated by alternating optimization algorithm. The SMSE algorithm can be decomposed into three steps.

Step 1. Supervised Local Neighborhood Embedding

Assume we have a multi-view feature set X of N subjects with M views, the labels of subject in X is represented as $Y = \{y_1, \dots, y_N\}$. The i^{th} view could be represented as $X^{(i)} = \{x_1^{(i)}, \dots, x_N^{(i)}\} \in \mathbb{R}^{V^{(i)} \times N}$, where $V^{(i)}$ is the dimension of the features of the i^{th} view. The neighborhood of $x_j^{(i)}$ is formed by itself and its K nearest neighbors, i.e., $X_j^{(i)} = \{x_j^{(i)}, x_{j_1}^{(i)}, \dots, x_{j_K}^{(i)}\} \in \mathbb{R}^{V^{(i)} \times (K+1)}$. The mapping of $X_j^{(i)}$ in the embedded feature space is $\hat{X}_j^{(i)} = \{\hat{x}_j^{(i)}, \hat{x}_{j_1}^{(i)}, \dots, \hat{x}_{j_K}^{(i)}\}$, and $\hat{X}_j^{(i)} \in \mathbb{R}^{d \times (K+1)}$, where d is the dimension of the aimed embedding, $d < V^{(i)}$. To preserve the local neighborhood information for all subjects, we need to solve the local optimization as in (1):

$$\arg \min_{\hat{X}, \alpha} \sum_{i=1}^M \alpha_i \sum_{j=1}^N \sum_{k=1}^K \left\| \hat{x}_j^{(i)} - \hat{x}_{j_k}^{(i)} \right\|^2 \cdot (\omega_{j_k}^{(i)}) \quad (1)$$

$$s.t. \quad \forall j_k, y_{j_k} = y_j$$

where $\hat{X}_j = \{\hat{x}_j^{(i)}\}$ are the local embeddings of neighborhoods of $x_j^{(i)}$ across all views, $\alpha = \{\alpha_i\}$ are a set of non-negative weights on individual views that indicate the contribution of each view, and $\omega_{j_k}^{(i)} \propto e^{-\|x_j^{(i)} - x_{j_k}^{(i)}\|^2}$ is the similarity between $x_j^{(i)}$ and $x_{j_k}^{(i)}$ in the original feature space.

Different from the original MSE, we enforce the local neighborhood, $X_j^{(i)}$, to be constructed by the neighbors from the same group as $x_j^{(i)}$. This constraint is used to model the

knowledge learning process, which preserves as much information as possible between subjects in the same group and summarizes the characteristics of the same group of features. Therefore the local neighborhood optimization in SMSE is indeed a ‘supervised’ process since the labels of the existing data are used in finding the local embeddings. With this constraint, the SMSE has a stronger intra-group merging effect in the local embeddings than the original MSE. Note that we set such constraint on the training data only when training a classifier, but not on the testing data.

Step 2. Global Coordinate Alignment

Each local neighborhood embedding, $\hat{X}_j^{(i)} \in \mathbb{R}^{d \times (K+1)}$, could be considered as a subset of the global embedding ($\hat{X} \in \mathbb{R}^{d \times N}$). The embeddings of individual views are consistent with each other in the global embedding. Therefore, Eq. (1) could be rewritten as:

$$\arg \min_{\hat{X}, \alpha} \sum_{i=1}^M \alpha_i^\gamma \sum_{j=1}^N \hat{X} L_j^{(i)} \hat{X}^T \quad (2)$$

$$s.t. \quad \hat{X} \hat{X}^T = I; \sum_{i=1}^M \alpha_i = 1, \alpha_i \geq 0$$

where $L_j^{(i)} \in \mathbb{R}^{N \times N}$ is the normalized Laplacian matrix for $\hat{X}_j^{(i)}$ which selects the K nearest neighbors and also sets diffusion coefficient to each neighbor according to $\omega_{jk}^{(i)}$ as:

$$L_j^{(i)}(j, j_k) = \begin{cases} 0 & \text{when } j = j_k \\ \frac{\omega_{jk}^{(i)}}{\sum_{k=1}^K \omega_{jk}^{(i)}} & \text{when } j \neq j_k \end{cases} \quad (3)$$

The constraint, $\hat{X} \hat{X}^T = I$, ensures the unique solution of \hat{X} . The scale parameter, γ , controls the scale of contributions of different views. When $\gamma = 1$, the solution to Eq. (2) is $\alpha_i = 1$ as the i^{th} view has the best performance. When $\gamma > 1$, \hat{X} will preserve the locality of each view simultaneously.

Step 3. Alternating Optimization

The solution of Step 2 is a non-convex optimization problem. We use the alternating optimization algorithm [16] to solve Eq. (2). We first initialize α as $\alpha = \{\frac{1}{M}, \dots, \frac{1}{M}\}$ and calculate \hat{X} as the smallest d eigenvectors of the L as:

$$L = \sum_{i=1}^M \alpha_i^\gamma \sum_{j=1}^N L_j^{(i)} \quad (4)$$

We then update α_i as:

$$\alpha_i = \frac{\text{tr}(\sum_{j=1}^N \hat{X} L_j^{(i)} \hat{X}^T)^{-\frac{1}{\gamma-1}}}{\sum_i \text{tr}(\sum_{j=1}^N \hat{X} L_j^{(i)} \hat{X}^T)^{-\frac{1}{\gamma-1}}} \quad (5)$$

We interactively update \hat{X} and α_i as in Eq. (4) and Eq. (5) until convergence.

In this study, the variables describing the data and the features were recalled as follows: N , the number of subjects, was 331; M , the number of views, was 4; $V^{(i)}$, the feature dimension of i^{th} view, was 83 for all the views; and y_j , the label of the j^{th} subject, belonged to one of the three labels $\{‘AD’, ‘MCI’, ‘CN’\}$. We further set the variables of the SMSE algorithm as follows: d , the dimension of the aimed embedding space, was set as 50; K , the number of nearest neighbors, was set as 10; and γ , the scale parameter for global embedding, was set as 3.

2.4 Performance Evaluation

We compared the performance of the proposed SMSE method to the benchmarks of single-view feature methods, the naive concatenated feature method, and the multi-view embedding methods, *i.e.*, DSE and MSE, which were evaluated using the same parameter settings as the SMSE. All these algorithms were evaluated using the same dataset as described in Section 2.1. We built a set of binary SVMs with the radial basis function (RBF) kernels as the classifiers and the classification accuracies of the three diagnosis groups were used to evaluate the performance of all the benchmark algorithms and the proposed SMSE algorithm. A leave-50%-out 5-fold cross-validation paradigm was adopted in performance evaluation. The optimal trade-off parameter (C) and the kernel parameter (γ) for SVM were estimated via grid-search. All SVM based cross-validations and performance evaluations were conducted using LIBSVM library [23].

3. RESULTS

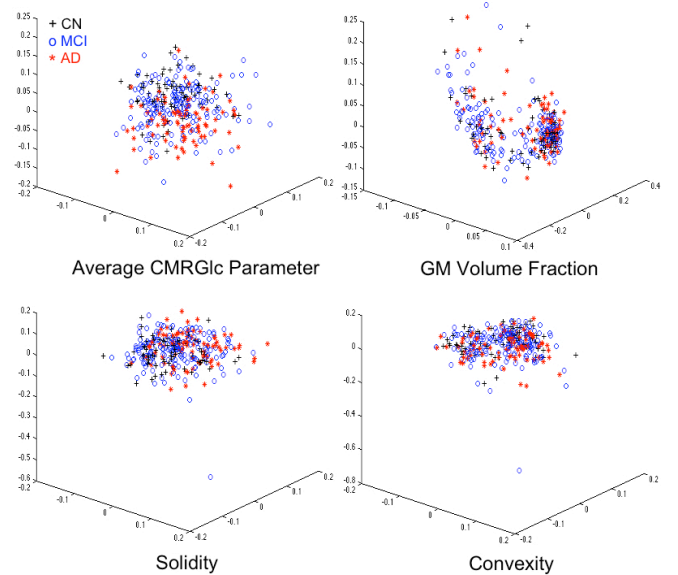


Figure 1. The embeddings of different views of features in 3D feature spaces generated using the kernel PCA algorithm.

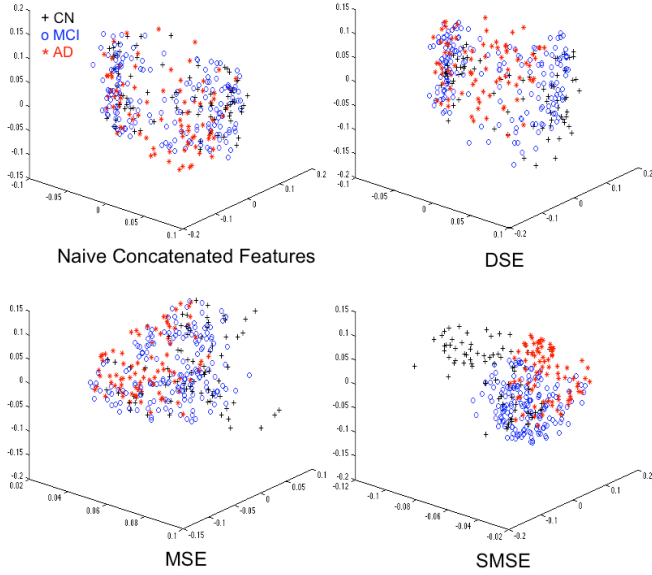


Figure 2. The embeddings of concatenated features, DSE, MSE and the proposed SMSE in 3D feature space. The kernel PCA algorithm was used to embed the concatenated features and the individual views of features in DSE.

Figure 1 shows the 3D embeddings of different views of features generated using the kernel PCA algorithm [24]. The three different classes in the single-view feature spaces were hardly separable. The different distributions of the four views of features suggested the complementary information among them. Figure 2 shows the embeddings derived from the naive concatenation, the unsupervised DSE and MSE, and the proposed SMSE methods. Both DSE and MSE could better separate the three groups of subjects than the naive concatenated features and the single-view features, as shown in Figure 1. SMSE demonstrates a strong intra-group merging tendency and makes the subjects more separable.

Table 1. The performance of all algorithms evaluated by the classification rates of the three diagnosis groups. $\langle d \rangle$ indicates the dimension of the features used in each algorithm.

Algorithm	Prediction Diagnosis	CN	MCI	AD
Best Single-View (Volume Fraction) $\langle d = 83 \rangle$	CN	38.13	52.39	8.94
	MCI	21.95	62.28	15.77
	AD	7.20	51.42	41.37
Concatenated Features $\langle d = 332 \rangle$	CN	32.80	64.38	2.82
	MCI	16.14	63.84	20.01
	AD	8.12	54.11	37.76
DSE $\langle d = 50 \rangle$	CN	41.67	55.72	2.61
	MCI	21.76	61.19	17.06
	AD	9.53	51.11	39.36
MSE $\langle d = 50 \rangle$	CN	44.29	52.81	2.90
	MCI	14.74	69.15	16.11
	AD	1.86	48.51	49.63
Proposed SMSE $\langle d = 50 \rangle$	CN	67.51	31.21	1.28
	MCI	14.79	84.62	0.59
	AD	4.68	29.34	65.98

Table 1 shows the classification rates of the diagnosis groups using different algorithms. We conducted the classification using the single-view features and found the best result achieved by single-view features was GM volume fractions, although other single-view features had comparable results. Concatenated features did not achieve better results than the single-view features, which implied the naive concatenated features failed to exploit the complementary information among single-view features. Both the unsupervised DSE and MSE methods achieved better results than the single-view based methods and the concatenation method, but the improvements of the unsupervised methods were limited. The proposed SMSE methods achieved the best performance among all the methods. The average classification rate of SMSE was 75.89%, compared to 58.23% of MSE, 51.10% of DSE, 49.80% of concatenation method.

4. DISCUSSIONS

In this study, we tested a set of spectral embedding methods, *i.e.*, LE, ISOMAP and locally linear embedding (LLE) [25] on the four types of single-view features. However, these methods did not work well because the K -nearest neighbors required by these methods had large variations due to the small group size (77-169 subjects) and the non-uniformly distributed features in the high dimensional feature space. The kernel PCA, on the contrary, was not affected by these disadvantages and preserved the overall smoothness of the original features. Therefore, we used kernel PCA for individual embeddings in DSE.

There are no direct solutions to estimating the SMSE parameters since the parameters are not guaranteed to be convex; therefore, we set the parameters, (γ, d, K) , as in Section 2.3, based on the pilot experiments. The effects of each parameter need to be investigated in details in the future work.

5. CONCLUSIONS

In this study, we proposed a supervised multi-view/multi-modal spectral embedding algorithm (SMSE) for neuroimaging classification based on the MSE framework. A supervised local neighborhood embedding method was proposed to replace the local optimization in unsupervised MSE. The proposed SMSE algorithm could embed the high dimensional multi-view neuroimaging features into a low dimensional feature space and preserve the optimal local information among different views. Overall the SMSE algorithm achieved better results than the unsupervised DSE and MSE methods, as well as the single-view based algorithms and the naive concatenation algorithm.

6. ACKNOWLEDGMENT

This study is supported in part by ARC and AADRF grants.

6. REFERENCES

- [1] S. Batty, J. Clark, T. Fryer and X. Gao, "Prototype system for semantic retrieval of neurological PET images," *The International Conference on Medical Imaging and Informatics (MIMI 2007), Aug 14-16, 2007, Beijing*, LNCS 4987, pp.179-188, 2008.
- [2] S. Liu, W. Cai, L. Wen, S. Eberl, M. Fulham and D. Feng, "Localized Functional Neuroimaging Retrieval using 3D Discrete Curvelet Transform," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2011)*, pp. 1877-1880, 2011.
- [3] S. Liu, W. Cai, L. Wen and D. Feng, "Volumetric Congruent Local Binary Patterns for 3D Neurological Image Retrieval," *IEEE International Conference on Image and Vision Computing New Zealand (IVCNZ 2011)*, 2011.
- [4] D. Unay, A. Ekin and R. S. Jasinski, "Local Structure-based Region of Interest Retrieval in Brain MR Images," *IEEE Trans. on Inf. Tech. in Biomed.*, vol. 14, issue. 4, pp.897-903, 2010.
- [5] S. Liu, W. Cai, L. Wen, S. Eberl, M. J. Fulham and D. Feng, "A Robust Volumetric Feature Extraction Approach for 3D Neuroimaging Retrieval," *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010)*, 2010.
- [6] S. Liu, W. Cai, L. Wen and D. Feng, "Neuroimaging Biomarker based Prediction of Alzheimer's Disease Severity with Optimized Graph Construction," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2013)*, 2013.
- [7] W. Cai, D. Feng and R. Fulton, "Content-Based Retrieval of Dynamic PET Functional Images," *IEEE Trans. on Information Tech. in Biomedicine*, vol. 4, no. 2, pp.152-158, 2000.
- [8] W. Cai, S. Liu, L. Wen, S. Eberl, M. Fulham and D. Feng, "3D Neurological Image Retrieval with Localized Pathology-Centric CMRGLC Patterns," *The IEEE 17th International Conference on Image Processing (ICIP 2010)*, pp. 3201-3204, 2010.
- [9] R. C. Petersen, G.E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos and E. Kokmen, "Mild Cognitive Impairment: Clinical Characterization and Outcome," *Arch. Neurol.*, vol. 56, pp. 303-308, 1999.
- [10] L. Shannon, A. J. Saykin, J. D. West, L. Shen, H. A. Firpi, B. C. McDonald and ADNI, "Baseline MRI Predictors for Conversion from MCI to Probable AD in the ADNI Cohort," *Current Alzheimer Research*, vol. 6, pp. 347-361, 2009.
- [11] J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo, V. A. Narayan and ADNI, "Sparse Learning and Stability Selection for Predicting MCI to AD Conversion using Baseline ADNI Data," *BioMed Central Neurology*, vol. 12, no. 46, 2012.
- [12] D. Zhang and D. Shen, "Multi-Modal Multi-Task Learning for Joint Prediction of Clinical Scores in Alzheimer's Disease," *MBIA 2011*, Nice, France, Oct. 1, 2012.
- [13] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," *Proc. Adv. Neural Inf. Process. Syst.*, pp. 585-591, 2002.
- [14] J. Tenenbaum, V. de Silva and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, Dec, 2000.
- [15] B. Long, P. S. Yu and Z. Zhang, "A General Model For Multiple View Unsupervised Learning," *Proc. 8th SIAM Int. Conf. Data Mining*, pp. 822-833, 2008.
- [16] T. Xia, D. Tao, T. Mei and Y. Zhang, "Multiview Spectral Embedding," *IEEE Tran. Syst. Man. Cybernetics*, vol. 40, no. 6, pp. 1438-1446, 2010.
- [17] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *J. Magnetic Resonance Imaging*, vol. 27, issue 4, pp. 685 – 691, 2008.
- [18] W. J. Jagust, D. Bandy, K. Chen, N. L. Foster, S. M. Landau, C. A. Mathis, *et al.*, "The Alzheimer's Disease Neuroimaging Initiative Positron Emission Tomography Core," *Alzheimer's & Dementia*, vol. 6, pp. 221-229, 2010.
- [19] M. Jenkinson, P. Bannister, J. M. Brady and S. M. Smith, "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images," *NeuroImage*, vol. 17, no. 2, pp. 825-841, 2002.
- [20] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, *et al.*, "A Probabilistic Atlas and Reference System for the Human Brain: International Consortium for Brain Mapping (ICBM)," *Phil. Trans. Royal Soc. B Biol. Sci.*, vol. 356, no. 1412, pp. 1293-1322, 2001.
- [21] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Harkens, *et al.*, "A Generic Framework for Non-rigid Registration based on Non-uniform Multi-level Free-form Deformations," *MICCAI2001*, pp. 573-581, 2001.
- [22] R. A. Heckemann, S. Keihaninejad, P. Aljabar, K. R. Gray, C. Nielsen, D. Rueckert, J. V. Hajnal, A. Hammers and ADNI, "Automatic Morphometry in Alzheimer's Disease and Mild Cognitive Impairment," *Neuroimage*, vol. 56, pp. 2024-2037, 2011.
- [23] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intel. Sys. Tech.*, vol.2, no.27, pp. 1-27, 2011.
- [24] B. Schölkopf, A. Smola and K. R. Müller, "Nonlinear Component Analysis As A Kernel Eigenvalue Problem" *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [25] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction By Locally Linearly Embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.