

# 알츠하이머 병의 검출을 위한 ML-SVM, PCA, VBM, GMM을 결합한 융합적 성능 비교

사우라르 알람, 권구락\*  
조선대학교 정보통신공학과

## Convergence performance comparison using combination of ML-SVM, PCA, VBM and GMM for detection of AD

Saurar Alam, Goo-Rak Kwon\*

Department of Information and Communication Engineering, Chosun University

**요약** 구조적 MRI 영상은 여러 단 변량과 다변량 방법을 위해 그레이 매터 (GM), 화이트 매터 (WM), 뇌척수액 (CSF) 세션화 과정을 하고 난후 형태계측학적 특징을 추출하기 위해 사용한다. 새로운 접근 방법은 매우 가벼운 알츠하이머 병에서 가벼운 알츠하이머병의 진단을 위해 적용된다. 간이정신상태검사에 따른 형태계측학적 특징과 가우시안 복합 모델 파라미터를 결합하여 정상인으로부터 알츠하이머 병 환자로 분류하는 방법을 제안한다. 결합한 특징은 주성분 분석 기법을 이용한 고차원의 저주를 제거한 후 다중 커널 SVM 분류기에 공급한다. 제안한 진단 방법의 실험적 결과는 90%이상의 특성도와 고민감도에 따라 다중 커널 SVM을 가진 층화 정확도가 96%까지 최대 산출한다.

• 주제어 : 알츠하이머 병, 의료IT융합, MMSE 점수, VBM 특징, GMM 파라미터, PCA, ML-SVM

**Abstract** Structural MRI(sMRI) imaging is used to extract morphometric features after Grey Matter (GM), White Matter (WM) for several univariate and multivariate method, and Cerebro-spinal Fluid (CSF) segmentation. A new approach is applied for the diagnosis of very mild to mild AD. We propose the classification method of Alzheimer disease patients from normal controls by combining morphometric features and Gaussian Mixture Models parameters along with MMSE (Mini Mental State Examination) score. The combined features are fed into Multi-kernel SVM classifier after getting rid of curse of dimensionality using principal component analysis. The experimenral results of the proposed diagnosis method yield up to 96% stratification accuracy with Multi-kernel SVM along with high sensitivity and specificity above 90%.

• Key Words : Alzheimer disease, Medical-IT convergence, MMSE Score, VBM features, GMM parameters, PCA, ML-SVM

\*Corresponding Author : 권구락(grkwon@chosun.ac.kr)

Received May 26, 2016

Revised June 1, 2016

Accepted August 8, 2016

Published August 31, 2016

## 1. Introduction

Several noninvasive efficient diagnosis methods can be applied like structural or functional magnetic resonance imaging (sMRI or fMRI), Single Photon Emission Computerized Tomography (SPECT), and Position Emission Tomography(PET) [1,2]. Several studies have unveiled that AD can be stratified from MCI and healthy controls(HC) by applying automatic or semi-automatic measurement of diversified priori brain Region of Interest (ROI). AD causes significant cerebral atrophy in several brain ROIs, especially the entorhinal cortex and hippocampus [3]. The regional volume of several ROIs has morphometric differences. The CAD tool has been designed [4] applying SPECT and PET image using NMF and SVM classifier for the classification of AD patients from healthy controls yielding up to 91% accuracy with high sensitivity and specificity rates (above 90%). M. López et al [5] developed PCA and Bayesian rule based tool for the diagnosis of AD patients from normal controls using SPECT and PET image dataset stratifying 98.3% accurately for SPECT and 88.6% accurately for PET dataset. The detection tool to identify CDR scale of entire 457 MR images has been developed using PCA and Artificial Neural Network, and classification accuracy achieving up to 89.92%. Multimodal approach [7] has been applied to discriminate AD and MCI patients from healthy controls using sMRI, FDG-PET, and CSF (cerebro-spinal fluid) as biomarkers. The stratification accuracy yielded up to 93.2% for AD vs HC and 76.4% for AD vs MCI. They have achieved accuracy of 86.5%, 72% differentiating AD from HC controls and MCI from HC respectively using best individual modality biomarker among them. Independent Component Analysis based AD classification [8] from HC controls has been proposed using both all gray matter images and whole brain images from the OASIS dataset.

The classification accuracy they got is  $73.7\% \pm 4.5\%$  of gray matter images from OASIS, and  $71.1\% \pm 4.5\%$  of gray matter images from the ADNI. The

discriminating accuracy of AD from MCI [9] has yielded up to 88.89% using SOM and PSO-SVM. Four different models of Artificial Neural Networks (ANNs) [10] have been applied to the same dataset, and 83% stratification accuracy has been achieved. Several univariate and multivariate method can be applied while Gray Matter (GM), White Matter (WM), and Cerebro-spinal Fluid (CSF) regions are being segmented. The preprocessing steps for VBM feature extraction follows Spatial Normalization, Segmentation, Smoothing, and Statistical Parametric Mapping. In this method, PCA based VBM and GMM feature along with MMSE score, and Multi-Kernel Support Vector Machine (ML-SVM) with train and test strategy is applied to classify new unknown subject.

## 2. System model and methods

The sMR images used here are spatially normalized into standard template with the help of SPM8 toolbox. Then images are segmented into GM, WM, and CSF. Then next steps follow smoothing of GM images and voxel-wise statistical tests to extract features. The feature along with GMM components and MMSE score are fed into multi-kernel SVM after applying two sample t-test and PCA for dimensionality reduction.

### 1.1 Overview of the experimental data

All sMRI images are collected from OASIS cross-sectional dataset. Ninety-eight women MRI data has been selected from there. Among them, 49 patients are diagnosed with very mild to mild AD and rest 49 patients are non-demented controls. The age, education, socio economic status, CDR, and MMSE score details of those are depicted in Table 1. The CDR details is summarized in Table 2.

### 1.2 Preprocessing and VBM feature extraction

The different steps of feature extraction process are depicted in Fig 1. The feature extraction process is

illustrated elaborately by Darya Chyzhyk et al [11]. VBM toolbox has been used for preprocessing. The OASIS cross sectional MR images have been already averaged, registered, resampled with 1-mm isotropic image in atlas space and bias field corrected there.

The images are realigned with template image, spatially normalized, segmented to get GM components, and then spatially smoothed with FWHM of the Gaussian kernel to 10mm isotropic prior to analyze voxel base statistics. A GM mask has been formed with the mean of GM segmentation volumes of the 98 subjects. The binary mask was generated by using thresholding operation on the average GM segmented volumes consisting of all voxels with probability greater than 0.1. The statistical General Linear Model (GLM) has been performed to analyze the groups mapped to very mild to mild AD patients and Normal controls using two-sample t-test. After setting different function in SPM toolbox and extracting features from VBM detected clusters, supervised learning algorithm are trained. The feature includes mean and standard deviation of detected clusters, and high dimensional vector having in all GM segmentation values of voxel location of all cluster.

<Table 1> Summary of subject demographics and dementia status

	Very mild to mild AD	Normal
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

<Table 2> Summary of a five-point scale, Clinical Dementia Rating (CDR)

CDR	Status
0.5	Very mild dementia
1	Mild
2	Moderate
3	Severe

### 1.3 Gaussian mixture model

Gaussian mixture model is a histogram probabilistic model, and can also be considered as an estimation of probability density. Image histogram reflects the frequency appears in a certain gray value. Given a grayscale image with N pixels divided into S different regions, so it depicts S peaks in the histogram. Thus, the image can be expressed by these S statistical models. Optimum thresholding for image segmentation problem is achieved by estimating the parameters of GMM. GMM based features provides promise accuracy to distinguish between target and ghost/clutter regions [12].

GMM is a parametric probability density function represented as a weighted sum of R Gaussian component densities,

$$pdf(x|\lambda) = \sum w_i g(x|m_i, \sum_i), \quad (1)$$

where,  $x$  is the N dimensional continuous valued data vector,  $w_i$  is the mixture weight, and  $g$  is the component Gaussian densities.

Each component density is evaluated by the N-variate Gaussian function,

$$g(x|m_i, \sum_i) = \frac{1}{(2\pi)^{N/2} |\sum_i|^{1/2}} \times \exp\left(-\frac{1}{2}(x - m_i)^T (\sum_i^{-1})(x - m_i)\right), \quad (2)$$

where,  $m_i$  is the average of a vector,  $T$  is the transpose, and  $\sum_i$  is the covariance matrix.

Using mean vectors, covariance matrices, and mixture weights from all component densities, the GMM was parameterized. These parameters are denoted as,

$$\lambda = \{w_i, m_i, \sum_i\}. \quad (3)$$

The variance ( $V_i$ ) of components is extracted from

the diagonal of the covariance matrix. The feature vector from three components of GMM,

$$R = \{w_1 \dots w_3, \mu_1 \dots \mu_3, v_1 \dots v_3\}, \quad (4)$$

where,  $w$ ,  $\mu$ , and  $v$  are the weight, average, and variance of GMM components.

#### 1.4 PCA and MK-SVM Classifier

In this paper, a noble approach is applied combining VBM feature and GMM parameters along with MMSE score, reducing the curse of dimensionality by using PCA, differentiating very mild AD to mild AD from healthy controls with Multi-kernel SVM as shown in Fig 2.

##### 1.4.1 Principal component analysis

In this method, PCA is used to reduce the dimensionality of higher dimensional featured after performing two sample t-tests. The PCA is summarized as

- Finding mean of the data matrix and zero mean matrix.
- Constructing covariance matrix.
- Getting the eigenvalue and the eigenvector.
- Projecting the data matrix with eigenvectors corresponding to highest to lowest eigenvalues.

##### 1.4.2 Multi-Kernel SVM

Support Vector Machine is supervised learning model, and it performs efficiently for binary classification of both linearly separable and non-separable data. Kernel trick is applied to make linearly non separable data to separable. The algorithm is about to get optimum hyperplane between two classes which has optimum distance from support vectors of both the classes. While it needs to take decision for new data point, classifier does that based on the hyperplane.

The kernel mapping function,  $\Phi: R^d \rightarrow R^f (f > d)$ , translates the feature data to

higher dimensional space so that it becomes linearly separable.

The linear decision boundary is

$$f(x) = w^T \phi(x) + w_0. \quad (5)$$

Let consider the training set  $\{(x_i, y_i)\}_{i=1}^n$ , where  $x_i \in R^d$  is training sample and its corresponding class level is  $y_i \in \{-1, +1\}$ , for SVM with L1 soft margin regularization can be solved with the primal problem,

$$\begin{aligned} \min_{w, w_0} & \frac{1}{2} \|w\|^2 + T \sum_{i=1}^n \psi_i, \\ \text{s.t.} & y_i (< w, \phi(x_i) > + w_0 \geq 1 - \psi_i, \forall i, \\ & \xi \geq 0, \end{aligned} \quad (6)$$

where  $T$  is the trade-off parameter of training error and margin,  $\psi_{1 \times n}$  is slack vector having non-zero elements.

The Lagrange dual space optimization solution is

$$\begin{aligned} \min_{\alpha} & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j), \\ \text{s.t.} & 0 \leq \alpha_i \leq T, \forall i; \sum_i \alpha_i y_i = 0, \end{aligned} \quad (7)$$

where,  $k$  is the kernel matrix  $\alpha_i, \alpha_j$  are the Lagrange multipliers.

Multi-kernel approach combining various kernels such as polynomial, Gaussian, sigmoid kernel can also be applied instead of using single kernel. Convex combination of multiple kernels is considered where each kernel has its kernel weight based on their discriminative power.

$$\begin{aligned} k(x_i, x_j) &= \sum_m \beta_m k_m(x_i, x_j), \\ \sum_m \beta_m &= 1, \beta \geq 0, \end{aligned} \quad (8)$$

where,  $\beta_m$  is the weight of m-th kernel.

Multi kernel learning Support Vector Machine can be solved in primal problem,

$$\begin{aligned} \min_{w, \beta, w_0} & \frac{1}{2} \frac{1}{\beta_m} \|w_m\|_2^2 + T \sum_{i=1}^n \psi_i, \\ \text{s.t.} & y_i (< w^T \sum_l \phi(x_i) > + w_0) \geq 1 - \psi_i, \forall i, \\ & \sum_m \beta_m = 1, \\ & \psi \geq 0, \beta \geq 0. \end{aligned} \quad (9)$$

The solution of optimization problem in Lagrange dual space,

$$\begin{aligned} \min_{\alpha} & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m \beta_m k_m(x_i, x_j), \\ \text{s.t.} & 0 \leq \alpha_i \leq T, \forall i; \\ & \sum_m \beta_m = 1, \beta \geq 0, \end{aligned} \quad (10)$$

where,  $\beta_m$  is the m-th kernel weight,  $\alpha_i, \alpha_j$  are the Lagrange multipliers.

The convex kernel SVM solution is solved as a dual problem and formulated as

$$\begin{aligned} \underset{\alpha}{\text{minimize}} & \frac{1}{2} \vec{\alpha}^T Y K Y \vec{\alpha} - \vec{\alpha}^T \vec{1}, \\ \text{subject to} & (Y \vec{\alpha})^T \vec{1} = 0, \\ & 0 < \alpha_i < T, i = 1, \dots, N \end{aligned} \quad (11)$$

where,  $Y = \text{diag}(y_1, \dots, y_N)$ , K is the kernel matrix.

To integrate multiple kernels instead of using single kernel [13], [14], [15], the program solution can be defined with L2 regularizations as

$$\begin{aligned} \underset{t, \alpha}{\text{minimize}} & \frac{1}{2} t - \vec{\alpha}^T \vec{1}, \\ \text{subject to} & (Y \vec{\alpha})^T \vec{1} = 0, \\ & 0 < \alpha_i < T, i = 1, \dots, N, t \leq \|\lambda\|_2 \end{aligned} \quad (12)$$

where,  $\vec{\lambda} = \{\vec{\alpha}^T Y K Y \vec{\alpha}, \dots, \vec{\alpha}^T Y K Y \vec{\alpha}\}$ ,  $\vec{\lambda} \in R^d$ . t is a variable to minimize.

The decision function is

$$f(x) = \text{sign}(\sum_m \beta_m k_m(x_i, x_j) + w_0). \quad (13)$$

In our proposed, polynomial, Gaussian, and sigmoid kernel are integrated. All the convex problem of kernel learning and multi-kernel learning is solved by second order cone programming (SOCP) using CVX and Sedumi package [16] for optimizing over symmetric cones. The kernel weights are calculated in dual space, and it provides importance of each kernel to distinguish AD from HC, or AD from MCI. RBF kernel SVM applied here is solved using LibSVM package [15].

### 3. PCA and MK-SVM Classifier result

To train MK-SVM for differentiating very mild to mild AD from normal controls, training samples are selected and tested the performance using selected testing samples. For drastically reducing the dimension of feature, PCA is applied on it. The supervised learning of GM segmented image features and GMM parameters along with clinical measurement (MMSE score) is used for diagnosis of AD patients. Multi-Kernel SVM is applied, and the accuracy and efficiency are measured with MMSE score or without the MMSE score. I have come across that when MMSE score is incorporated with GM segmented image features and GMM parameters, the classification accuracy, sensitivity, and specificity gets higher as portrayed in Table 3.

### 4. Conclusions

A diagnosis method stratifying normal controls from very mild to mild AD is unveiled here. The noble method is enhanced by performing PCA which reduces the dimensionality of feature space significantly and selecting the most three principal components in terms of ability to differentiate for training and testing Multi kernel SVM for the

diagnosis. We obtained almost 96% accuracy with high sensitivity and specificity which outperforms the classification accuracy of 83% obtained by Savio A. et al [10],  $73.7\% \pm 4.5\%$  by WenluYanga et al [8], and 89.92% by R. Mahmood et al [6].

<Table 3> Classification result of very mild to mild AD from HC

Method	Accuracy	Sensitivity	Specificity
PCA+Linear SVM	83.33	83.33	83.33
PCA+ Multi-Kernel SVM	84.17	85.17	86.17
MMSE score+PCA+ Linear SVM	95.83	100.00	91.67
MMSE score+PCA+ Multi Kernel SVM	96.33	97.17	95.50

## ACKNOWLEDGMENTS

This study was supported by research funds from Chosun University, 2015. And this research was supported by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C7A1046050).

## REFERENCES

- [1] Ramesh Kumar Lama and G.-R. Kwon, "Multiresolution Non-Local Means Filtering for Image Denoising," *The Journal of Korean Institute of Next Generation Computing*, Vol. 11, No. 5, pp. 17-23, Oct. 2015.
- [2] Jeongjin Lee, Che Hwan Seo, Juneseuk Shin, and Yeong-Gil Shin, "Accurate Liver Vascular Structure Analysis in Abdominal CT Images," *The Journal of Korean Institute of Next Generation Computing*, Vol. 11, No. 2, pp. 41-48, Apr. 2015.
- [3] M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, and O. Colliot, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI, Hippocampus," Vol. 19, pp. 579-587, 2009.
- [4] P. Padilla, M. Lopez, 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 Transactions on Medical Imaging*, Vol. 31, Issue 2, pp. 207 - 216, Feb. 2012.
- [5] M. López, J. Ramírez, J. M. Górriz, D. Salas Gonzalez, I. Álvarez, F. Segovia, and C. G. Puntonet, "Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules," *Electronics Letters*, Vol. 45, Issue 8, pp. 389 - 391, Apr. 2009.
- [6] R. Mahmood, and B. Ghimire, "Automatic detection and classification of Alzheimer's Disease from MRI scans using principal component analysis and artificial neural networks," *Systems, Signals and Image Processing (IWSSIP)*, 20th IEEE International Conference, pp. 133 - 137, 2013.
- [7] Wenlu Yanga, Ronald L. M. Luib, Jia-Hong Gaoc, Tony F. Chand, Shing-Tung Yaub, Reisa A. Sperlinge, and Xudong Huang f, "Independent Component Analysis-Based Classification of Alzheimer's Disease MRI Data," *Journal of Alzheimer's Disease* 24, pp. 775-783, 2011.
- [8] Shih-Ting Yang, Jiann-Der Lee, Tzyh-Chyang Chang, Chung-Hsien Huang, Jiun-Jie Wang, Wen-Chuin Hsu, Hsiao-Lung Chan, Yau-Yau Wai, and Kuan-YiLi, "Discrimination between Alzheimer's Disease and Mild Cognitive Impairment Using SOM and PSO-SVM," *Computational and Mathematical Methods in Medicine*, Vol. 2013, Article ID 253670, 10 pages, 2013.
- [9] Savio A, and Garcia-Sebastian M, "Classification results of artificial neural networks for Alzheimer's disease detection," *Lect. Notes Comput. Sci.*, Vol. 5788, pp. 641-648, 2009.
- [10] Darya Chyzhyk, Alexandre Savio, "Feature

extraction from structural MRI images based on VBM: data from OASIS database”, University of the Basque Country, Internal research publication.

[11] V. Kilaru, M. Amin, F. Ahmad, P. Sévigny, and D. DiFilippo, “Gaussian mixture modeling approach for stationary human identification in through-the-wall radar imagery,” *Journal of Electronic Imaging*, Vol. 24, No. 1, Article ID 013028, 2015.

[12] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, M.L. Jordan, Learning the Kernel Matrix with Semidefinite Programming, *Journal of Machine Learning Research* 5 pp. 27-72, 2005.

[13] GRG. Lanckriet, T. De Bie, N. Cristianini, M.L. Jordan, and WS. Noble, “A statistical framework for genomic data fusion,” *Bioinformatics*, Vol. 20, pp. 2626-2635, 2004.

[14] F.R. Bach, G.R.G. Lanckriet, and M.L. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” *Proceedings of 21st International Conference of Machine Learning*, 2004.

[15] M. Grant and S. Boyd, CVX, “Matlab software for disciplined convex programming,” version 2.0 beta, 2013.

[16] C.Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” 2001.

권 구 락(Goo-Rak Kwon)

[정회원]



· Feb. 2007: PhD degrees in mechatronic engineering, Korea university  
 · Mar. 2008-present, Associate professor, Department of information and communication Engineering, Chosun university

<관심분야> : Research interests: Multimedia signal processing, Image communication, Bio-infomatics

저자소개

알람 사루알(Saurar Alam)



· 2011 BS degree in Department of Computer & Communication Engineering, International Islamic University Chittagong  
 · 2014 MS degree in Department of Information & Communication Engineering, Chosun University

<관심분야> : Research interests: Medical Image segmentation and processing, Pattern recognition, Feature extraction