



Published in final edited form as:

Mach Learn Med Imaging. 2015 ; 9352: 255–262. doi:10.1007/978-3-319-24888-2_31.

Multi-view Classification for Identification of Alzheimer's Disease

Xiaofeng Zhu¹, Heung-Il Suk², Yonghua Zhu³, Kim-Han Thung¹, Guorong Wu¹, and Dinggang Shen¹

Dinggang Shen: dgshen@med.unc.edu

¹Department of Radiology and BRIC, University of North Carolina, Chapel Hill, USA

²Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

³School of Computer, Electronics and Information, Guangxi University, Nanning, China

Abstract

In this paper, we propose a multi-view learning method using Magnetic Resonance Imaging (MRI) data for Alzheimer's Disease (AD) diagnosis. Specifically, we extract both Region-Of-Interest (ROI) features and Histograms of Oriented Gradient (HOG) features from each MRI image, and then propose mapping HOG features onto the space of ROI features to make them comparable and to impose high intra-class similarity with low inter-class similarity. Finally, both mapped HOG features and original ROI features are input to the support vector machine for AD diagnosis. The purpose of mapping HOG features onto the space of ROI features is to provide complementary information so that features from different views can *not only* be comparable (*i.e.*, homogeneous) *but also* be interpretable. For example, ROI features are robust to noise, but lack of reflecting small or subtle changes, while HOG features are diverse but less robust to noise. The proposed multi-view learning method is designed to learn the transformation between two spaces and to separate the classes under the supervision of class labels. The experimental results on the MRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset show that the proposed multi-view method helps enhance disease status identification performance, outperforming both baseline methods and state-of-the-art methods.

1 Introduction

Alzheimer's Disease (AD) is the most popular form of dementia among the elderly population. It is estimated that there are around 90 million AD patients in the world, with the number of AD patients expected to reach 300 million by 2050 [8,12]. In this regard, it is very interesting and important to find an accurate biomarker for the diagnosis of AD and its prodromal stage, *i.e.*, Mild Cognitive Impairment (MCI). For the past few decades, neuroimaging has been widely used to investigate AD-related pathologies in the spectrum between cognitive normal and AD [7,17], where various machine learning techniques have been designed for the analysis of complex patterns in neuroimaging data, as well as identification of a subject's clinical status. For example, Cuingnet *et al.* embedded a graph-

based regularization operator into Support Vector Machine (SVM) for the identification of AD [2], while Wang *et al.* designed a sparse Bayesian multitask learning model to adaptively investigate the dependence of AD subjects, for improving the AD diagnosis performance [10].

Since multi-modality data (including Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and CerebroSpinal Fluid (CSF) biomarkers) are often acquired in applications and have been shown to provide complementary information for AD diagnosis [4,5,11,13,16], a great number of research use multi-modality data for AD diagnosis and obtain significant performance improvements, compared to the methods that use a single modality data [9,15,19]. For example, Zhang *et al.* designed an approach that conducts AD diagnosis by directly concatenating features of multiple modalities of data including MRI data, PET data, and CSF data, as their method outperformed other methods with individual modality data, such as MRI data or PET data [13,18]. However, to the best of our knowledge, very few previous works have focused on the identification of AD with multi-view or visual features of neuroimaging data.

In this paper, we propose a new multi-view learning method using multiple representations of MRI images for AD diagnosis, via the following three stages: 1) *Image processing*. We extract both Region-Of-Interest (ROI) features and 3-dimensional Histograms of Oriented Gradient (HOG) [6] features from given MRI images. 2) *Multi-view learning*. A new multi-view learning method is designed to map HOG features onto the space of ROI features, by ensuring high similarity for samples with the same label, while low similarity for samples with different labels. This makes classes well separated. 3) *AD classification*. Both the mapped HOG features and the original ROI features are fed into a SVM classifier to identify AD.

Compared to conventional methods (*e.g.*, the multi-modality method [13] and the single-view method [9]) for AD diagnosis, this work has the following contributions.

- We extract both HOG features and ROI features from only MRI images to form multi-view features, rather than conventional multi-modality methods using both MRI images and PET images [13]. That is, multi-modality methods need to pay additional for PET images, whereas no additional payments are required for our method. In practice, the ANDI dataset provides more MRI images (*e.g.*, more than 800) than PET images (*e.g.*, only about 400), and has been indicated that less training data can easily result in under-fitting [9].
- Few studies focus on AD diagnosis via visual features, such as HOG, even though HOG features and ROI features can provide complementary information. It has been shown that ROI features, *e.g.*, the average of gray matter volume within a brain region, are robust to the noise but are less diverse for AD diagnosis [13]. In contrast, HOG features can output multiple bi-dimensional histograms for a brain region to reflect the change of blocks within a brain region, so HOG features are good at reflecting small or subtle changes within brain, though vulnerable to noises [6].

- Compared to learning common space among different views like Canonical Correlation Analysis (CCA) [14], the proposed method learns the mappings from the HOG feature space to the ROI feature space, with the guidance of learning high intra-class similarity and low inter-class similarity.

2 Approach

2.1 Notations

We denote matrices as boldface uppercase letters, vectors as boldface lower-case letters, and scalars as normal italic letters, respectively. For a matrix $\mathbf{X} = [x_{ij}]$, its i -th row and j -th column are denoted as $\mathbf{x}^i$ and $\mathbf{x}_j$, respectively. The Frobenius norm and the transpose

operator of a matrix $\mathbf{X}$ are denoted as $\|\mathbf{X}\|_F = \sqrt{\sum_i \|\mathbf{x}^i\|_2^2} = \sqrt{\sum_j \|\mathbf{x}_j\|_2^2} = \sqrt{\sum_{i,j} x_{ij}^2}$ and $\mathbf{X}^T$, respectively.

2.2 Image Processing

The ROI feature can be regarded as the global feature, as it is obtained by averaging gray matter tissue volume within a brain region. The ROI feature has been indicated to be robust to noise but very coarse in sense, such as the lack of reflecting small or subtle changes involved in brain diseases [9]. However, disease-related structural/functional changes may occur in multiple brain regions [13]. Therefore, the simple ROI representation may not effectively capture disease-related pathologies. In contrast, the HOG feature decomposes a 3D image into a grid of small squared 3D cells, where a bi-dimensional histogram of gradient along spatial and orientation bins is computed and then returns a descriptor for each cell [6]. The HOG feature considers the diversity of each cell and can be regarded as the local feature. In this work, we simultaneously extract ROI features (*i.e.*, global features) and HOG features (*i.e.*, local features) from an MRI image to form a multi-view representation for AD diagnosis. To do this, we use 830 MRI images (including 198 ADs, 403 MCIs and 229 Normal Controls (NC)) from ADNI database¹. Additionally, we select 124 progress MCIs (pMCI) and 118 stable MCIs (sMCI) from 403 MCIs to conduct the binary classification pMCI vs. sMCI².

More specifically, we first preprocessed the MRI images by performing spatial distortion, skull-stripping, and cerebellum removal, sequentially, and then segmented the MRI images into gray matter, white matter, and cerebrospinal fluid. Furthermore, the MRI images were parcellated into 93 ROIs based on a Jacob template, by non-rigid brain registration. We finally computed the gray matter tissue volumes of the ROIs as the ROI features.

Given the MRI images separated by 93 ROIs, we extracted the HOG features for each ROI. Specifically, we first down-sampled the original MRI images, *i.e.*, from $256 \times 256 \times 256$ to $64 \times 64 \times 64$, followed by partitioning the whole brain into 93 ROIs, which is the same as partitioning the original brain image to extract the ROI features. We dilated each ROI with 3 voxels to achieve a soft boundary among ROIs. Following the method in [6], we set the

¹Please refer to '<http://adni.loni.usc.edu/>' for up-to-date information.

²In ADNI, these numbers, *i.e.*, 124 and 118, of subjects were, respectively, marked as pMCI and sMCI among 403 MCI subjects.

number of orientation bins to 9, with each bin with 8 orientations to describe a descriptor by a 72-dimensional feature vector. We also set the size (in voxels) of the spatial bins and the size of the blocks, respectively, as 5 and 2, to extract 1728 descriptors from each ROI. Note that descriptor information was divided into overlapping blocks, each of which contained $2 \times 2 \times 2$ 3-dimensional cells. We further clustered descriptors of each ROI of all MRI images to form a 50-dimensional bag-of-words for each ROI.

Finally, we used a 93 dimensional ROI feature vector and a 4650 ($= 93 \times 50$) dimensional HOG feature vector to obtain a multi-view representation of an MRI image.

2.3 Multi-view Learning

A conventional solution of multi-view learning is to search for a common space among different views. For example, Canonical Correlation Analysis (CCA) was designed to search a common space among views in which the diversity among all views was minimized [14]. However, recent studies indicate that such a common space obtained by a symmetric transformation (*i.e.*, the same rotation and scaling to all views) cannot separate classes particularly well [3,14]. To address this, we design a new multi-view learning method to transform HOG features into the ROI feature space by ensuring that the HOG-ROI feature pairs of the same label have high similarity (*i.e.*, high intra-similarity), while those of different labels have low similarity (*i.e.*, low inter-similarity).

Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_n\}$ denote, respectively, HOG features and ROI features of n samples, where $\mathbf{x}_i \in \mathbb{R}^{d_x}$, $\mathbf{y}_i \in \mathbb{R}^{d_y}$, and d_x and d_y are the dimensionalities of the HOG and ROI features, respectively. We, then, learn a transformation matrix $\mathbf{W} \in \mathbb{R}^{d_x \times d_y}$ from the HOG feature space to the ROI feature space (or equivalently a transformation matrix $\mathbf{W}^T$ from the ROI feature space to the HOG feature space). We first define an inner product similarity function between any sample pair, *i.e.*, $\mathbf{x}_i \in \mathbf{X}$ and $\mathbf{y}_j \in \mathbf{Y}$, as follows

$$\text{sim}(\mathbf{x}_i, \mathbf{y}_j) = \mathbf{x}_i^T \mathbf{W} \mathbf{y}_j, \quad i, j = 1, \dots, n. \quad (1)$$

In finding a transformation matrix $\mathbf{W}$, we also expect Eq. (1) to have high intra-similarity for samples of the same class, but low inter-similarity for samples of different classes under the supervision of the class labels. In this regard, we formulate the following cost function for a given sample pair $(\mathbf{x}_i, \mathbf{y}_j)$, $i, j = 1, \dots, n$:

$$c(\mathbf{x}_i, \mathbf{y}_j) = \begin{cases} \max(0, \text{sim}(\mathbf{x}_i, \mathbf{y}_j) - \mu), & \text{if } l(i) = l(j) \\ \max(0, \nu - \text{sim}(\mathbf{x}_i, \mathbf{y}_j)), & \text{otherwise} \end{cases} \quad (2)$$

where $l(i)$ (or $l(j)$) is the label of the HOG features of the i -th sample (or the label of the ROI features of the j -th sample), μ and ν are upper and lower bound parameters to guarantee the constraint, *i.e.*, the largest value of $c(\mathbf{x}_i, \mathbf{y}_j)$, $i, j = 1, \dots, n$, for the sample pair $(\mathbf{x}_i, \mathbf{y}_j)$ with the same class label and the smallest value $c(\mathbf{x}_i, \mathbf{y}_j)$ for the sample pair $(\mathbf{x}_i, \mathbf{y}_j)$ with different class labels. Finally, we define a loss function over all sample pairs as follows

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}) = \sum_{i,j=1}^n \{c(\mathbf{x}_i, \mathbf{y}_j)\}^2 \quad (3)$$

To avoid over-fitting, we add a Frobenius norm into Eq. (3) to get the final objective function as follows:

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{X}, \mathbf{Y}) + \lambda \|\mathbf{W}\|_F^2 \quad (4)$$

where $\lambda > 0$ is a tuning parameter. Due to the convexity of both the cost function and the regularization term in Eq. (4), the optimization of Eq. (4) has a global optimum. We employ an alternating projection method based on Bregman's algorithm [1] to optimize Eq. (4). Specifically, the Bregman's method updates the transformation matrix $\mathbf{W}$ with respect to a single constraint in Eq. (2) of a sample pair for each time, which can be easily scaled to large-scale problems and fast convergence in practice.

2.4 AD Classification

After obtaining the transformation matrix $\mathbf{W}$, we concatenate the transformed HOG features $\mathbf{x}_i^T \mathbf{W}$ with the original ROI features $\mathbf{y}_i$ to form a new representation

$\mathbf{z}_i = [\mathbf{x}_i^T \mathbf{W}, \mathbf{y}_i^T]^T \in \mathbb{R}^{2d_y}$. Naturally, we can also directly concatenate the original HOG features with the original ROI features to form another new representation

$\mathbf{z}_i = [\mathbf{x}_i^T, \mathbf{y}_i^T]^T \in \mathbb{R}^{(d_x+d_y)}$. We can unify these two kinds of representation as

$\mathbf{z}_i = [f(\mathbf{x}_i), \mathbf{y}_i^T]^T$, where

$$f(\mathbf{x}_i) = \begin{cases} \mathbf{x}_i^T \mathbf{W}, & \text{Linearly mapping HOG to ROI} \\ \mathbf{x}_i^T, & \text{Concatenating HOG with ROI} \end{cases} \quad (5)$$

In this work, we call the former case (*i.e.*, $f(\mathbf{x}_i) = \mathbf{x}_i^T \mathbf{W}$) as the Single-direction Mapping Multi-view Learning (SMML for short) method and the latter case (*i.e.*, $f(\mathbf{x}_i) = \mathbf{x}_i$) as the Directly Concatenating Multi-view Learning (DCML for short) method. Note that SMML transfers HOG features into the space of the ROI features, via a linear transformation matrix (*i.e.*, $\mathbf{W}$), while DCML does that via an identity matrix. We, then, use a linear SVM as a classifier since it has been shown that SVM does not encounter the issue of curse of the dimensionality.

3 Experimental Results and Discussion

We conducted various classification tasks on the ADNI dataset (www.adni-info.org) to justify the effectiveness of the proposed method.

3.1 Experimental Setting

In our experiments, we considered three binary classification tasks, *i.e.*, AD vs. NC, MCI vs. NC, and pMCI vs. sMCI, to compare our DCML and SMML with the baseline methods (*i.e.*,

the SVM classification via the HOG features (HOG for short) and the SVM classification via the ROI features (ROI for short), respectively) and the state-of-the-art methods (*i.e.*, CCA [14] and Multiple Instance Learning method on MRI images (MIL for short) [9]). Among the competing methods, single-view methods include HOG, ROI, and MIL, respectively, while multi-view methods include CCA and the proposed DCML and SMML.

For each binary classification task, we followed the steps of (1) extracting HOG and ROI features; (2) finding a transformation matrix $\mathbf{W}$; (3) conducting SVM learning; and (4) evaluating the performance with classification accuracy.

We used a 10-fold cross-validation method in our experiments. In each fold, we conducted 5-fold inner cross-validation for model parameter selection by a line search method on the parameters with the predefined range, such as $\lambda \in \{10^{-5}, \dots, 10^5\}$ in the LIBSVM toolbox. Regarding the upper and lower bounds in Eq. (2), we set them as $\mu = 1$ and $\nu = -1$. The parameters that resulted in the best performance in the inner cross-validation were finally used in testing. We repeated the process 10 times to avoid any possible bias occurring in data partitioning for cross-validation. The final performance was reported by averaging the repeated cross-validation results.

3.2 Performance and Discussion

Table 1 shows the performance of all other competing methods. The proposed SMML achieved the best performances for all three binary classification tasks, followed by CCA, DCML, MIL, ROI and HOG, respectively. For example, SMML achieved improvements of 9.86%, 6.40%, and 6.08%, respectively, on AD vs. NC, MCI vs. NC, and pMCI vs. sMCI, compared to the worst method of all competing methods, *i.e.*, HOG method, and improved by 1.08%, 2.41%, and 2.61%, respectively, on AD vs. NC, MCI vs. NC, and pMCI vs. sMCI, compared to the CCA that achieved the best performance among competing methods.

MIL outperformed all other single-view methods, such as HOG and ROI. MIL is a patch-based method and extracts ROI features within each patch, as it is diverse and also robust, compared to either ROI or HOG. However, our proposed methods outperformed MIL. For example, compared to MIL, our SMML improved by 1.61%, 5.18%, and 8.29%, while our DCML increased by 0.29%, 2.34%, and 4.87%, on AD vs. NC, MCI vs. NC, and pMCI vs. sMCI, respectively. Additionally, HOG outperformed ROI for two out of three classification tasks, such as MCI vs. NC and pMCI vs. sMCI. This indicated that visual feature (*i.e.*, HOG) is useful for AD diagnosis. Besides, multi-view methods (such as CCA, DCML, and SMML) were better than any single-view method (such as HOG, ROI, and MIL). This showed that HOG features and ROI features provide complementary information. However, the proposed SMML still outperformed CCA since our method simultaneously achieved high intra-class similarity and low inter-class similarity during the estimation of transformation, compared to CCA results.

4 Conclusion

In this paper, we proposed a new multi-view learning method to identify AD using MRI images. The experimental results on the ADNI dataset showed that the proposed method

outperformed the state-of-the-art methods for AD diagnosis, as our multi-view representation provides complementary information by extracting both global features (*i.e.*, ROI features) and local features (*i.e.*, HOG features) from MRI images and further imposing high intra-class similarity and low inter-class similarity during feature mapping.

Acknowledgments

This work was supported in part by NIH grants (EB006733, EB008374, EB009634, MH100217, AG041721, AG042599). Xiaofeng Zhu was supported in part by the National Natural Science Foundation of China under grant 61263035. Heung-Il Suk was supported in part by ICT R&D program of MSIP/IITP [B0101-15-0307, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)].

References

1. Censor, Y. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press; 1997.
2. Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, Chupin M, Benali H, Colliot O. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage*. 2011; 56(2):766–781. [PubMed: 20542124]
3. Harel M, Mannor S. Learning from multiple outlooks. *ICML*. 2011:401–408.
4. Jin Y, Shi Y, Zhan L, Gutman BA, de Zubicaray GI, McMahon KL, Wright MJ, Toga AW, Thompson PM. Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. *NeuroImage*. 2014; 100:75–90. [PubMed: 24821529]
5. Li, J.; Jin, Y.; Shi, Y.; Dinov, ID.; Wang, DJ.; Toga, AW.; Thompson, PM. Voxelwise spectral diffusional connectivity and its applications to alzheimer's disease and intelligence prediction. In: Mori, K.; Sakuma, I.; Sato, Y.; Barillot, C.; Navab, N., editors. *MICCAI 2013, Part I*. LNCS. Vol. 8149. Springer; Heidelberg: 2013. p. 655–662.
6. Sanroma G, Wu G, Gao Y, Shen D. Learning to rank atlases for multiple-atlas segmentation. *IEEE Transactions Medical Imaging*. 2014; 33(10):1939–1953.
7. Suk HI, Lee SW, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function*. 2013; 220(2):841–859. [PubMed: 24363140]
8. Thung K, Wee C, Yap P, Shen D. Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *NeuroImage*. 2014; 91:386–400. [PubMed: 24480301]
9. Tong T, Wolz R, Gao Q, Guerrero R, Hajnal JV, Rueckert D. Multiple instance learning for classification of dementia in brain MRI. *Medical Image Analysis*. 2014; 18(5):808–818. [PubMed: 24858570]
10. Wan J, Zhang Z, Yan J, Li T, Rao BD, Fang S, Kim S, Risacher SL, Saykin AJ, Shen L. Sparse bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in alzheimer's disease. *CVPR*. 2012:940–947.
11. Zhan L, Jahanshad N, Jin Y, Toga AW, McMahon K, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Brain network efficiency and topology depend on the fiber tracking method: 11 tractography algorithms compared in 536 subjects. *ISBI*. 2013:1134–1137.
12. Zhan L, Zhou J, Wang Y, Jin Y, Jahanshad N, Prasad G, Nir TM, Leonardo CD, Ye J, Thompson PM. Comparison of 9 tractography algorithms for detecting abnormal structural brain networks in alzheimers disease. *Frontiers in Aging Neuroscience*. 2015; 7(48):401–408.
13. Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*. 2012; 59(2):895–907. [PubMed: 21992749]
14. Zhu X, Huang Z, Shen HT, Cheng J, Xu C. Dimensionality reduction by mixed kernel canonical correlation analysis. *Pattern Recognition*. 2012; 45(8):3003–3016.

15. Zhu X, Li X, Zhang S. Block-row sparse multiview multilabel learning for image classification. *IEEE Transactions on Cybernetics*. 2015
16. Zhu X, Suk HI, Shen D. Matrix-similarity based loss function and feature selection for alzheimer's disease diagnosis. *CVPR*. 2014:3089–3096. [PubMed: 26379415]
17. Zhu X, Suk HI, Shen D. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage*. 2014; 100:91–105. [PubMed: 24911377]
18. Zhu, X.; Suk, H-I.; Shen, D. A novel multi-relation regularization method for regression and classification in AD diagnosis. In: Golland, P.; Hata, N.; Barillot, C.; Hornegger, J.; Howe, R., editors. *MICCAI 2014, Part III*. LNCS. Vol. 8675. Springer; Heidelberg: 2014. p. 401-408.
19. Zhu X, Zhang L, Huang Z. A sparse embedding and least variance encoding approach to hashing. *IEEE Transactions on Image Processing*. 2014; 23(9):3737–3750. [PubMed: 24968174]

Table 1

Comparison of the classification accuracy (mean $\pm$ standard deviation) of all methods at different classification tasks.

Method	AD vs. NC	MCI vs. NC	pMCI vs. sMCI
HOG	0.8145 $\pm$ 0.0957	0.7167 $\pm$ 0.2088	0.6946 $\pm$ 2.5119
ROI	0.8969 $\pm$ 0.0951	0.7136 $\pm$ 0.1899	0.6638 $\pm$ 2.2902
MIL	0.8970 $\pm$ 0.0871	0.7289 $\pm$ 0.1249	0.6725 $\pm$ 1.4298
CCA	0.9023 $\pm$ 0.0838	0.7566 $\pm$ 0.1152	0.7293 $\pm$ 1.3333
DCML	0.8999 $\pm$ 0.0987	0.7523 $\pm$ 0.0991	0.7212 $\pm$ 1.2468
SMML	0.9131 $\pm$ 0.0629	0.7807 $\pm$ 0.0961	0.7554 $\pm$ 1.1972