NEUROIMAGING BIOMARKER BASED PREDICTION OF ALZHEIMER'S DISEASE SEVERITY WITH OPTIMIZED GRAPH CONSTRUCTION

Sidong Liu[1], *Weidong Cai*[1], *Lingfeng Wen*[2], *Dagan Feng*[1]

[1] Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Australia
[2] Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

ABSTRACT

The prediction of Alzheimer's disease (AD) severity is very important in AD diagnosis and patient care, especially for patients at early stage when clinical intervention is most effective and no irreversible damages have been formed to brains. To achieve accurate diagnosis of AD and identify the subjects who have higher risk to convert to AD, we proposed an AD severity prediction method based on the neuroimaging predictors evaluated by the region-wise atrophy patterns. The proposed method introduced a global cost function that encodes the empirical conversion rates for subjects at different progression stages from normal aging through mild cogitative impairment (MCI) to AD, based on the classic graph cut algorithm. Experimental results on ADNI baseline dataset of 758 subjects validated the efficacy of the proposed method.

Index Terms— Alzheimer's disease, neuroimaging, prediction

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurological degenerative disease, whose dementia symptoms gradually degraded over a number of years. Mild cognitive impairment (MCI) represents the transitional state between AD and cogitative normal aging subjects (NS) with a high conversion rate to AD [1]. For example, about 40% MCI patients from the ADNI (see Section 2.1) baseline cohort converted to AD within two years, compared to 3.9% for normal aging subjects during the same period. Predicting AD severity and identifying the patients with higher risk of AD conversion is very important, because some intervention or prevention measures can be applied to such patients to avoid irreversible damages in brain.

Neuroimaging is a fundamental component for AD diagnosis, and also an essential indicator in assessment of therapy and monitoring disease progression. Many neurodegenerative pattern classification methods have been proposed for diagnosis of AD and MCI with various neuroimaging data [2-8]. Most existing studies for AD have been simplified into two-class classification problems, *i.e.*, AD vs. NS [2-7] and/or MCI vs. NS [5-8]. However, the diagnosis of AD is indeed a multi-class classification problem that the whole spectrum of AD as the pathology evolves from normal aging through MCI converters (*c*MCI) and non-converters (*nc*MCI) to dementia need to be investigated. In other words, we need to identify the AD, *c*MCI and *nc*MCI subjects in a single setting. Prediction of AD severity is very challenging since multiple classes have more interferes with each other than two classes.

There are several studies on differentiating subjects at different severity stages. Fan *et al.* [9] designed an ordinal ranking system for detecting MCI (*c*MCI and *nc*MCI) and AD subjects based on multimodal neuroimaging and cerebrospinal fluid (CSF) biomarkers. Ye *et al.* [10] recently proposed a sparse learning based predictor selection method for predicting MCI conversions using various types of data including magnetic resonance imaging (MRI), demographic, genetic and cognitive measures. Shannon *et al.* [11] conducted a similar study purely based on MRI data. These studies worked in the same fashion. The neurodegenerative patterns were modeled by a set of predictors from neuroimaging data, usually MRI or positron emission tomography (PET), and integrated with others predictors, *e.g.*, CSF and clinical assessment scores. The selected predictors were then used to train the classifiers, *e.g.*, support vector machine (SVM) for future classification.

We believe that such workflow can be optimized by encoding certain prior pathology information into the prediction model, such as the empirical conversion rates of MCI and NS subjects. Therefore, in this study, we proposed a new graph construction algorithm for predicting AD severity based on the neuroimaging predictors, which was selected by the evaluation of information gain. We designed a new global-cost function to encode the conversion rates at different stages in addition to the data-cost and smooth-cost functions of the original graph cut algorithm. The advantage of the proposed method is that it does not make predictor-independence assumptions required by other methods in order to ensure tractable inference. Furthermore, the smooth-cost and new global-cost functions can regularize the model to ensure optimal prediction performance.

2. METHODS

2.1. Data Acquisition

The imaging data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu) [12]. Totally 816 subjects were selected from the ADNI baseline cohort, and a T1 weighted 1.5T or 3T MRI image was acquired from each subject. The subjects for study were selected using the following recruitment criteria. Firstly, we excluded 20 subjects with the multiple conversions or reversions, except only one subject who converted from NS through MCI to AD in three years and did not revert to MCI or NS ever again. Secondly, we excluded 21 normal subjects that converted to MCI, because we were not sure whether they would eventually convert to AD or remain as MCI. Finally, if the MCI subjects converted to AD in half year to 3 years from the first scan, they were thought as the MCI converters, denoted by '*c*MCI'. Other non-converser MCI subjects were referred to as '*nc*MCI'. The normal subjects and AD patients were labeled as 'NS' and 'AD', respectively.

All 'raw' MRI data were converted to the ADNI format following the ADNI MRI image correction protocol [12]. We then nonlinearly registered the MRI images to the ICBM_152 template [13] using the Image Registration Toolkit (IRTK) [14]. 17 images were excluded from the MRI dataset because the intolerable image distortions were found by visual check. Therefore, totally 758 participants were recruited based on the above criteria and visual check, including 180 AD subjects, 160 *c*MCI subjects, 214 *nc*MCI subjects and 204 normal aging subjects.

2.2. Feature Extraction and Selection

We mapped 83 brain structures in the template space using the multi-atlas propagation with enhanced registration (MAPER) approach [15] on each registered MRI image and then computed the grey matter (GM) volume of the brain regions as the candidate predictors, $v \in \mathbb{R}^{83}$. We further normalized the GM volume of each brain region by the volume of the brain mask as a fraction of the whole brain.

The two-tailed t-tests were carried out on each brain region and the *p-value* for each predictor was used to evaluate the significance. We applied t-test on four group-pairs, AD-NS, *c*MCI-NS, *nc*MCI-NS, and *c*MCI-*nc*MCI. Note that the *p-values* in this context were not used for the usual purpose of hypothesis testing, but for assessing the discriminant power of the predictor. Usually a cut-off value was set up to select the predictors, e.g.: $\alpha = 0.05\ or\ 0.01$. We take a further step to select the predictors by calculating the performance gain of different predictors as:

$$PGR(i) = 1 - \exp\left(-\frac{(\Pr(i) - \Pr(i-1))^2}{2\delta^2}\right) \quad (1)$$

$PGR(i)$ is the performance gain response of the i^{th} predictor sorted in ascending order based on its *p-value*. We selected these predictors whose performance gain is greater than smallest precision unit δ ($\delta = 0.1\%$ in this study).

2.3. Optimized Graph Construction

Graph cut algorithm [16] is a widely used vision method with many applications in image segmentation, motion estimation, image restoration and stereo matching. In a standard graph cut model, the data-cost that encodes the data point property and the smooth-cost that regularizes the close neighbors to encourage similar labels were defined as:

$$E(f|G) = \overbrace{\sum_{v \in V} D_v(f_v)}^{\text{Data Cost}} + \overbrace{\sum_{uv \in E} S_{uv}(f_u, f_v)}^{\text{Smooth Cost}} \quad (2)$$

where f is the label function on the graph $G = \langle V, E \rangle$, V is the vertex set and E is the edge set. Therefore, these vision problems could be naturally transformed to an energy minimization problem, *i.e.*:

$$\hat{f} = \arg\min_f E(f|G) \quad (3)$$

Inspired by the study of Song *et al.* [17], we considered the feature space as a n-dimension grid space, $\mathbb{R}^n$. The subjects of different groups were distributed in the space and the prediction of each subject depended on the subject's characteristic (encoded by the data-cost function) and the relationships with its neighbors (encoded by the smooth-cost function). Therefore our AD severity prediction problem, or equivalently, the multi-class classification problem, could be transformed to an energy minimization problem.

We assumed the distribution for AD, *c*MCI, *nc*MCI and NS groups were all multivariate Gaussians in $\mathbb{R}^n$. We estimated the mean (μ_*) and covariance matrix (σ_*) for each Gaussian using the expectation maximization algorithm and encoded the data-cost for each subject as negative log likelihood as:

$$D_v(f_v) = -\ln P(f_v|\mu_*, \sigma_*) \quad (4)$$

where $*$ indicated one of the four groups, and $P(f_v|\mu_*, \sigma_*)$ was the multivariate probability distribution function (PDF) given μ_* and σ_*. We further modeled the smooth-cost as:

$$S_{uv}(f_u, f_v) \propto \exp\left(-\frac{\phi(f_u, f_v)^2}{2 * \tau^2}\right) \cdot \frac{1}{\zeta(u, v)} \quad (5)$$

where $\phi(f_u, f_v)$ was the normalized mutual information between f_u and f_v, τ was the standard deviation of all uv in E, and $\zeta(u, v)$ was the normalized Euclidean distance between u and v in the feature space. This function penalized a lot for connected subjects when the mutual information difference is less than τ; and meanwhile

penalized less if the subjects were far away from each other. We used a 5-neighbour system in this study.

Different groups of subjects have different AD conversion rates. Therefore, in addition to the original cost functions in graph cut model, we introduced a new global cost function, which encoded such prior knowledge as a global-cost function, *i.e.*:

$$H(\ell_*) \propto -\ln(\vartheta(\ell_*)^2) \cdot \frac{1}{\ln((2\pi e)^n|\sigma_*|)} \quad (6)$$

where $\vartheta(\ell_*)$ is the empirical conversion rate for subjects with label $\ell_* \in \mathcal{L}, \mathcal{L} := \{\ell_{AD}, \ell_{cMCI}, \ell_{ncMCI}, \ell_{NS}\}$. The function $\ln((2\pi e)^n|\sigma_*|)$ was the entropy of $P(\mu_*, \sigma_*)$ as in Eq. (4), and $|\sigma_*|$ was the determinant of the covariance matrix σ_*. We set $\vartheta(\ell_*)$ for *c*MCI and NS as their annual conversion rate, 15% and 1.6% respectively [1]. For *nc*MCI and AD groups, we set 0.1% for both of them to mimic the stability of them. We further tuned the cost function for different groups with their entropies. The MCI groups were expected to have greater entropy than AD and NS groups, because they had more expending distributions than AD and NS. Therefore, this function penalized less if a MCI subject was mislabeled as AD or NS, whereas penalized more if an AD or NS subject was mislabeled as MCI.

Therefore, the AD severity prediction problem was transformed to minimization of the energy as:

$$\hat{f} = \arg\min_f (E(f|G) + \sum_{\ell \in \mathcal{L}} H(\ell|G)) \quad (7)$$

In this study, we solved the problem in Eq. (7) using the GCO_V3.0 library [18]. We used the $\alpha-$expansion algorithm [16] because it had strictly optimality bound and required a metric for defining S_{uv} to ensure optimal performance.

2.4. Performance Evaluation

In the feature selection step, the predictor's response to the performance gain was evaluated by Eq. (1) based on the SVM classification rate. Same set of selected predictors was used for both the proposed method and comparison method.

We compared our proposed method with the well-known SVM. The leave-10%-out 10-fold cross validation was adopted for both methods. During each iteration of the cross validation, 90% subjects were evaluated by AD severity prediction in the proposed method. In the contrast, the SVM model used 90% subjects to train the classifiers, and the other 10% subjects were used for evaluation. In this study, we designed a set of binary SVM classifiers with the radial basis function (RBF) kernel. The optimal trade-off parameter and the kernel parameter (C, γ) for SVM were estimated via grid-search. All SVM based feature selection, parameter estimation and classification were conducted using LIBSVM library [19].

3. RESULTS

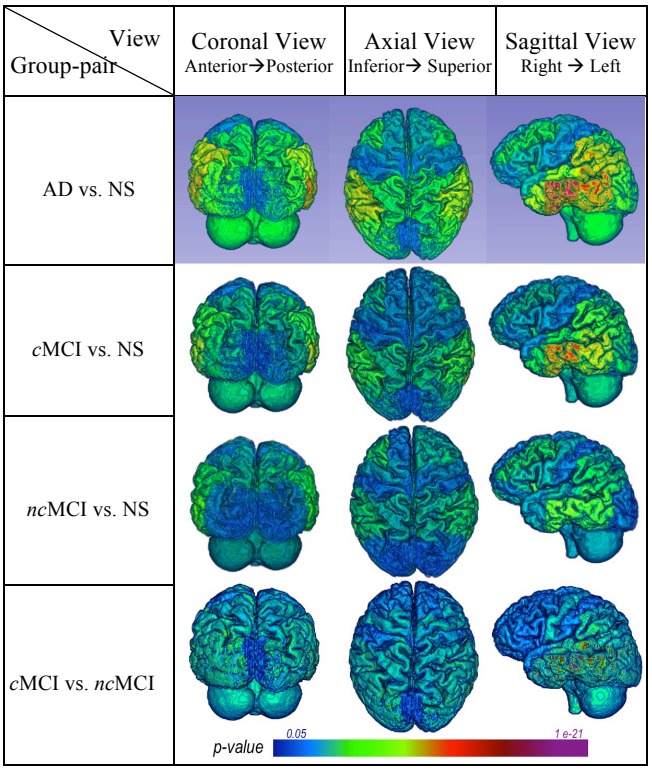

Figure 1 The heat maps of pair-wise atrophy pattern comparisons. The images were generated by 3D Slicer (V4.1) [20].

Figure 1 illustrates the atrophy heat maps of 83 regions among different group-pairs. The temporal lobe was the most distinctive predictor among different diagnostic groups. The AD-NS group-pair showed greater contrast than the other group-pairs. The atrophy differences between *n*MCI and *nc*MCI were the smallest, suggesting the differentiation between *n*MCI and *nc*MCI would be the most difficult.

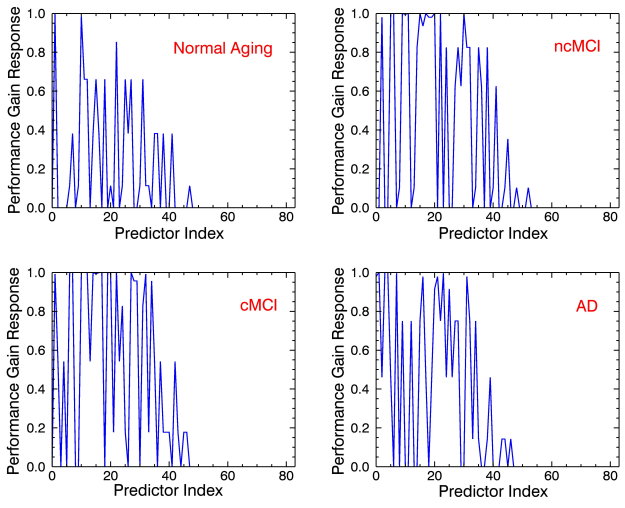

Figure 2 The performance gain responses for individual predictors in the four diagnostic groups.

Figure 2 shows the performance gain responses for individual predictors in prediction of the four diagnostic groups. The predictors' responses corresponded well to their *p-values*, although several predictors with significant *p*-values did not contribute to the prediction performance. The union of the selected predictors in individual groups formed a 52-predictor set that was consequently used in later prediction.

Table 1 The prediction performances (%) of the proposed method compared to the optimized SVM and original graph cut algorithm.

Algorithm	Prediction / Diagnosis	NS	*nc*MCI	*c*MCI	AD
Optimized SVM $(C, \gamma) = (0.95, 0.21)$	NS	81.9	13.2	1.5	3.4
	*nc*MCI	40.7	39.7	2.8	16.8
	*c*MCI	20.6	28.1	18.1	33.1
	AD	10.0	15.6	7.2	67.2
Original GC	NS	51.5	48.0	0.5	0.0
	*nc*MCI	0.0	99.5	0.5	0.0
	*c*MCI	1.3	38.1	60.0	0.0
	AD	2.2	17.2	1.1	79.4
Optimized GC (Proposed)	NS	66.2	32.8	0.1	0.0
	*nc*MCI	0.0	99.1	0.9	0.0
	cMCI	2.5	19.4	72.5	1.9
	AD	3.3	10.6	8.3	77.8

Table 1 shows the prediction performances of these four groups achieved by each method. Each row of the table shows the prediction accuracy in percentage of testing subjects of different diagnostic groups. The prediction of *c*MCI was the most difficult task compared to other groups and most of the misclassifications occur between *c*MCI and *nc*MCI patients. Optimized SVM achieved highest performance on NS (81.9%), but worst performance on *c*MCI (18.1%) and *nc*MCI (39.7%). Both original and optimized graph cut (GC) methods had better results than SVM, except for NS. GC methods had a strong tendency to predict NS as *nc*MCI. This is because *nc*MCI was sparsely distributed in the feature space and when a smooth function was used, marginal NS, *c*MCI and AD subjects tend to shift to *nc*MCI. The newly introduced global-cost function reduced the impact of the smooth-cost function. As a result, the optimized GC method retained the good performance on *nc*MCI and AD, and meanwhile improved the performance on *c*MCI and NS by 14.7% and 12.5%, respectively.

4. CONCLUSION

In this study, we investigated the entire spectrum of AD and proposed a new graph construction algorithm for predicting AD severity. A new global-cost function that encodes the empirical AD conversion rates for different diagnostic groups was proposed and integrated to the classic graph cut algorithm. This global-cost function lessened the negative impact of the smooth-cost function in the original algorithm. Significant improvement was achieved in the prediction of *nc*MCI, *c*MCI and AD as compared to the optimized SVM and the original graph cut algorithm.

5. ACKNOWLEDGEMENT

This work was supported in part by ARC grants.

6. REFERENCES

[1] Y. Fan and ADNI, "Ordinal Ranking for Detecting Mild Cognitive Impairment and Alzheimer's Disease based on Multimodal Neuroimages and CSF Biomarkers," *MBIA2011*, pp. 44-51, 2011.

[2] S. Kloppel, *et al.*, "Automatic Classification of MR Scans in Alzheimer's Disease," *Brain*, vol. 131, pp. 681-689, 2008.

[3] K. R. Gray, P. Aljabar, *et al.*, "Random Forest-based Manifold Learning for Classification of Imaging Data in Dementia" *MICCAI Workshop on MLMI*, pp. 154-161, 2011.

[4] S. Liu, W. Cai, L. Wen, S. Eberl, M. Fulham, and D. Feng, "Localized Functional Neuroimaging Retrieval using 3D Discrete Curvelet Transform", *ISBI2011*, pp1877-1880, 2011.

[5] Y. Wang, Y. Fan, *et al.*, "High Dimensional Pattern Regression using Machine Learning: From Medical Images to Continuous Variables," *Neuroimaging*, vol. 50, pp. 1519-1535, 2010.

[6] J. E. Iglesias, J. Jiang, *et al.*, "Classification of Alzheimer's Disease Using a Self-Smoothing Operator," *MICCAI2011*, Part III, LNCS 6893, pp. 58-65, 2011.

[7] T. Liu, *et al.*, "Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net," *Proc. 1st Int. Conf. Multimodal Brain Image Analysis (MBIA2011)* pp. 27-34, 2011.

[8] Y. Fan, *et al.*, "Spatial Patterns of Brain Atrophy in MCI Patients, Identified via High Dimensional Pattern Classification, Predict Subsequent Cognitive Decline," *Neuroimaging*, v39, pp.1731-43, 2008.

[9] R. C. Petersen, G.E. Smith, *et al.*, "Mild Cognitive Impairment: Clinical Characterization and Outcome," *Arch. Neurol.*, vol. 56, pp. 303-308, 1999.

[10] J. Ye, M. Farnum, *et al.*, "Sparse Learning and Stability Selection for Predicting MCI to AD Conversion using Baseline ADNI Data," *BioMed Central Neurology*, vol. 12, no. 46, 2012.

[11] L. Shannon, A. J. Saykin, *et al.*, "Baseline MRI Predictors for Conversion from MCI to Probable AD in the ADNI Cohort," *Current Alzheimer Research*, vol. 6, pp. 347-361, 2009.

[12] C. R. Jack, M. A. Bernstein, *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods," *J. Magnetic Resonance Imaging,* vol. 27, issue 4, pp. 685 – 691, 2008.

[13] J. Mazziotta, *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.

[14] J. A. Schnabel, *et al*., "A Generic Framework for Non-rigid Registration based on Non-uniform Multi-level Free-form Deformations," *MICCAI2001*, pp. 573-581, 2001.

[15] R. A. Heckemann, S. Keihaninejad, *et al.*, "Automatic Morphometry in Alzheimer's Disease and Mild Cognitive Impairment," *Neuroimage*, vol. 56, pp. 2024-2037, 2011.

[16] Y. Boykov, O. Veksler and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts", *IEEE Trans. Pattern Rec. Machine Learning*, vol. 23, no. 11, pp. 1222-1239, 2001.

[17] Y. Song, W. Cai, Y. Wang and D. D. Feng, "Location Classification of Lung Nodules with Optimized Graph Construction," *ISBI2012*, vol.1, pp. 1439-1442, 2012.

[18] A. Delong, *et al.*, "Fast Approximate Energy Minimization with Label Costs," *Int. J. Comp. Vision*, vol. 96, no. 1, pp.1-27, 2012.

[19] C. C. Chang, and C. J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. on Intel. Sys. Tech*, vol. 2, no. 27, pp. 1:27, 2011.

[20] S. Pieper, B. Lorensen, *et al.*, "The NA-MIC Kit: ITK, VTK, Pipelines, Grids and 3D Slicer as an Open Platform for the Medical Image Computing Community," *ISBI2006*, vol.1, pp. 698-701, 2006.