BOOSTING CLASSIFICATION ACCURACY OF DIFFUSION MRI DERIVED BRAIN NETWORKS FOR THE SUBTYPES OF MILD COGNITIVE IMPAIRMENT USING HIGHER ORDER SINGULAR VALUE DECOMPOSITION

L. Zhan[1, 2], Y. Liu[3*], J. Zhou[3], J. Ye[4,5], P.M. Thompson[2]*

1. Dept. of Neurology, University of California, Los Angeles, CA 90095, USA
2. Imaging Genetics Center, Keck School of Medicine,
University of Southern California, Marina Del Rey, CA 90292, USA
3. Dept. of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287, USA
4. Dept. of Computational Medicine and Bioinformatics,
5. Dept. of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48105, USA;
* Equal Contribution.

ABSTRACT

Mild cognitive impairment (MCI) is an intermediate stage between normal aging and Alzheimer's disease (AD), and around 10-15% of people with MCI develop AD each year. More recently, MCI has been further subdivided into early and late stages, and there is interest in identifying sensitive brain imaging biomarkers that help to differentiate stages of MCI. Here, we focused on anatomical brain networks computed from diffusion MRI and proposed a new feature extraction and classification framework based on higher order singular value decomposition and sparse logistic regression. In tests on publicly available data from the Alzheimer's Disease Neuroimaging Initiative, our proposed framework showed promise in detecting brain network differences that help in classifying early versus late MCI.

***Index Terms*—** Mild Cognitive Impairment, diffusion MRI, brain network, high order SVD, classification

1. INTRODUCTION

Mild cognitive impairment (MCI) involves the onset and evolution of cognitive impairments beyond those expected based on the age and education of the individual, but which are not significant enough to interfere with their daily activities [1]. It is also considered to be a transitional stage between normal aging and dementia, as every year, ~10% to 15% of people with MCI progress to probable AD [2] However, not all people with MCI deteriorate cognitively and some even get better. More accurate accurate classification of MCI subtype is crucial for disease detection and treatment evaluation.

Mild cognitive impairment is a "clinical" diagnosis representing a doctor's best professional judgment about the reason for a person's symptoms. There are currently no tests or procedures to demonstrate conclusively that a person has MCI. Impairment on cognitive tests is usually defined as performance at least 1.5 standard deviations (SD) below the age-, sex- and education-adjusted mean on a standardized test. In an attempt to define an even earlier point in time for disease detection, the recent extensions of the North American ADNI project (termed, "ADNI-GO" and ADNI-2) introduced the distinction of early versus late MCI. Late MCI (LMCI) refers to the original definition (performance of 1.5 SD below the normative mean), whereas in early MCI (EMCI), impairment is defined as performance in the range 1.0-1.5 SD below the normative mean on a standardized tests [3].

Recently studies of brain connectivity – including anatomical connectivity assessed with diffusion MRI – have attracted increasing attention. Reconstruction and modeling of brain networks provides an alternative way to investigate brain diseases on a holistic scale. More and more studies show that brain network properties are altered in certain disorders, such as bipolar illness [4, 5], body dysmorphic disorder [6], and even HIV/AIDS [7]. Even, brain networks and their features depend to some extent on the choice of fiber tracking algorithm used to infer the trajectories of pathways in the brain [8-10]; also, dozens of tractography algorithms are now available [11-17] yielding visually very different brain networks. In this study, we adopted tensor-based fiber assignment by continuous tracking (FACT) algorithm [12] to compute the brain network. Tensor-based FACT can yield false positive fibers which may add noise to computed network properties, but FACT is still one of the most widely used tractography algorithms, and it is simple and flexible. Here we propose a novel framework for network classification, with the goal of improving diagnostic classification based on networks. We also set out to show how this new framework could be applied to noisy

networks (such as those derived from FACT) and used for differentiating EMCI from LMCI.

2. METHODS

Figure 1 illustrates the basic idea of our proposed framework for brain network classification using higher order singular value decomposition (HO-SVD) and sparse logistic regression (Sparse LG). Its two component techniques are explained below.

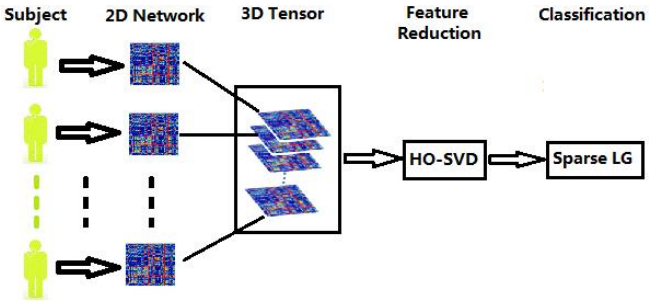

Figure 1. Basic workflow for network feature reduction and classification. The approach involves computing brain networks as connectivity matrices, but then stacking them up, across subjects, as a 3D tensor and performing feature reduction and sparse methods for classification.

2.1 HO-SVD

In machine learning and data mining, SVD is a powerful tool for dimension reduction. The SVD of a matrix $X \in \mathbb{R}^{n\times m}$ is given by $X = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{m\times m}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{n\times m}$ is a rectangular diagonal matrix. The diagonal entries of Σ, known as singular values, are non-negative and assumed to be in descending order.

The higher order SVD (HO-SVD) is one common generalization of SVD from matrices to tensors [18]. In HO-SVD, a tensor $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ is decomposed as

$$\mathcal{X} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \ldots \times_N U^{(N)},$$

in which

(1) $U^{(k)} \in \mathbb{R}^{I_k\times I_k}, k = 1, \ldots, N$ are orthogonal matrices where the ith column of $U^{(k)}$ is the ith k-mode singular vector.

(2) $\mathcal{S} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ is the core tensor which is of the same size as $\mathcal{X}$, and has the following properties:

• For any $1 \le k \le N$, let $\mathcal{S}_{i_k}$ and $\mathcal{S}_{j_k}$ be the subtensors obtained by fixing the kth index to i_k and j_k, $1 \le i_k, j_k \le I_k$, then $\langle \mathcal{S}_{i_k}, \mathcal{S}_{j_k} \rangle = 0$ for $i_k \ne j_k$;

• For $1 \le k \le N$,

$$\| \mathcal{S}_{i_k=1} \| \ge \| \mathcal{S}_{i_k=2} \| \ge \cdots \ge \| \mathcal{S}_{i_k=I_k} \| \ge 0;$$

The Frobenius-norms $\| \mathcal{S}_{i_k=i} \|, 1 \le i \le I_k$ are the k-mode singular values.

The kth mode singular matrix $U^{(k)}$ can be obtained as the left singular matrix of the kth mode unfolding matrix of tensor $\mathcal{X}$. After obtaining all N singular matrices $U^{(1)} \ldots U^{(N)}$, the core tensor $\mathcal{S}$ is given by

$$\mathcal{S} = \mathcal{X} \times_1 U^{(1)^T} \times_2 U^{(2)^T} \ldots \times_N U^{(N)^T}.$$

Inspired by the dimension reduction via SVD in the 2D case, we propose to reduce the feature dimensions, using higher order SVD. Similar to the matrix case, the ordering assumption for tensor singular values suggests that most of the information contained in a tensor may be expressed by the first few "components". Let the first mode of data tensor $\mathcal{X}$ correspond to the sample size n (*i.e.*, $I_1 = n$) and the remaining modes correspond to feature dimensions. Then, by keeping the largest $R_1, \ldots, R_N$ singular values for each mode, a reduced tensor with size $n \times R_2 \times R_3 \times \ldots \times R_N$ can be obtained by

$$\tilde{\mathcal{X}} = \tilde{\mathcal{S}} \times_1 \tilde{U}^{(1)},$$

where $\tilde{\mathcal{S}} = \mathcal{X} \times_1 \tilde{U}^{(1)^T} \times_2 \tilde{U}^{(2)^T} \ldots \times_N \tilde{U}^{(N)^T}$ is a tensor with size $R_1 \times R_2 \times R_3 \times \ldots \times R_N$, $\tilde{U}^{(k)} \in \mathbb{R}^{I_k\times R_k}, 1 \le \mathrm{k} \le N$, $\tilde{\mathcal{S}}$ is attained by keeping the first $R_1, \ldots, R_N$. The proposed dimension reduction of the tensor is also analogous to principal components analysis [19] for a matrix input. Instead of the original tensor, we propose to use the reduced tensor $\tilde{\mathcal{X}}$ as the new input data for classification.

2.2 Sparse Logistic Regression

Let $x \in \mathbb{R}^m$ be a sample vector and $y \in \{-1, +1\}$ be a binary outcome. The logistic regression model is given by:

$$\mathrm{Prob}(y|x) = \frac{1}{1+\exp(-y(x^T w + c))},$$

where $w \in \mathbb{R}^m$ and $c \in \mathbb{R}$ are coefficients, and Prob(y|x) is the posterior probability. The empirical logistic loss is measured by the negative log-likelihood and the average logistic loss is given by

$$\mathcal{L}(w, c) = -\frac{1}{n}\log \prod_{i=1}^{n} \mathrm{Prob}(y_i|a_i) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i(x^T w + c))).$$

The unknown coefficients w and c can be computed by minimizing the logistic loss, and this involves a smooth convex optimization problem. However, when dimension m is far larger than the sample size n, solving the logistic regression problem is ill-posed, and the learned model may suffer from the over-fitting problem.

Sparse logistic regression embeds the feature selection into classification using the Lasso penalty [20] which results in a sparse solution for w. The sparse logistic regression problem is formulated as:

$$\min_{w,c} \mathcal{L}(w, c) + \lambda \| w \|_1,$$

where the l_1 norm of w, *i.e.*, $\| w \|_1$, is the Lasso penalty and $\lambda > 0$ is the regularization parameter that controls the sparsity of the solution.

3. EXPERIMENT DESIGN

3.1 Subject demographics and image acquisition

Data used in this study were obtained from ADNI2, the second stage of the Northern American Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu). In total, 112 subjects were analyzed; of these, there were 73 people with EMCI (mean age: 71.34±11.57, 47 males) and 39 with LMCI (mean age: 72.32±5.83, 24 males).

Each subject underwent whole-brain MRI scanning on 3-Tesla GE Medical Systems scanners. T1-weighted SPGR (spoiled gradient echo) sequences (256x256 matrix; voxel size = 1.2x1.0x1.0 mm^3; TI=400 ms; TR = 6.98 ms; TE = 2.85 ms; flip angle = 11°), were collected as well as diffusion-weighted images (DWI; 256x256 matrix; voxel size: 2.7x2.7x2.7 mm^3; scan time = 9 min; more imaging details may be found at http://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_GE_3T_22.0_T2.pdf). 46 separate images were acquired for each DWI scan: 5 T2-weighted images with no diffusion sensitization (b_0 images) and 41 diffusion-weighted images (b=1000 s/mm²). The DWI protocol for ADNI was chosen after a detailed evaluation of different protocols that could be performed in a reasonable amount of time; we reported results of these comparisons previously [21, 22]. All T1-weighted MR and DWI images were checked visually for quality assurance to exclude scans with excessive motion and/or artifacts; all scans were included.

3.2 Network Computation

Each subject's brain network was computed using the method described in [23]. In brief, each subject's DWI was preprocessed (such as corrected for eddy current distortion and motion as well as removal of non-brain tissue) using FSL toolbox (http://fsl.fmrib.ox.ac.uk/). Then whole brain tractography was computed using tensor-based fiber assignment by continuous tracking (FACT) algorithm [12] implemented in diffusion toolkit (http://trackvis.org/dtk/). 113 cortical and subcortical ROIs were defined using the Harvard Oxford Cortical and Subcortical probabilistic atlas [24]. For each pair of ROIs, the number of detected fibers connecting them was determined from the FACT tractography. A fiber was considered to connect two ROIs if it intersected both of them. This process was repeated for all ROI pairs, to compute a whole brain fiber connectivity matrix. This matrix is symmetric, by definition, and has a zero diagonal (no self-connections; please see [23] for details). To avoid possible computation bias in the later feature extraction and evaluation sections, we normalized each matrix by the maximum value in the matrix, as matrices derived from different subjects have different scales and ranges. This normalized network will serve as the feature for the classification.

3.3 Feature extraction and classification

Based on a total of 112 subjects' 113x113 normalized networks computed in the above section, we compared three feature extraction methods:

(1) Raw features: for each subject, the feature vector was constructed by stacking all the entries of the upper triangular matrix, since the brain matrix is symmetric. So each subject has 6328 (=113x112/2) features. Thus, we obtained the 112 (subjects) x 6328 (features) matrix as the input to Sparse LG.

(2) SVD: We first built the raw feature matrix and then center this matrix by subtracting the means for each column. We use the top 10 principal components as the input for Sparse LG.

(3) HO-SVD: We reduce the dimension of data tensor to 112x10x10 by keeping the largest 10 singular values for each mode. Then, we construct the feature vector for each subject by stacking the entries of the reduced data matrix. This constructed feature vector then serves as the input for Sparse LG.

After feature extraction, we normalize each feature column of the input data matrix by subtracting the mean value from it and dividing it by the standard deviation. Since the outcome distribution is highly imbalanced, we used under-sampling to mitigate the bias caused. For each training/testing procedure, 85% LMCIs and EMCIs were randomly drawn as the training set, while the remaining subjects were used for testing. We then ran Sparse LG on the training set and evaluate classification performance on the test set, with parameters selected by 5-fold cross-validation. We repeat the training/test procedure 20 times, and report the mean and standard deviations in classification performances including accuracy, sensitivity, specificity, area under the curve (AUC) and the balanced accuracy which is the mean of sensitivity and specificity. The Sparse LG model is implemented using the Sparse Learning with Efficient Projections package.[25] In our experiments, LMCI was designated as the positive class and EMCI was the negative class.

4. RESULTS AND DISCUSSIONS

There was no significant difference (P=0.53) in average age between the subjects with EMCI and LMCI. The gender ratio (the number of males over the total number of subjects) was also comparable (0.64 for EMCI and 0.62 for LMCI). **Figure 2** shows the mean brain network for EMCI and LMCI. Although EMCI and LMCI have very similar visual patterns in the mean network, the mean network difference plot indicates there are indeed differences in some matrix cell values. When we performed paired T-test on each cell of the network between EMCI and LMCI and the false discovery rate method (FDR) was used to correct for multiple comparisons, our results indicated there were no

element-wise differences between the group mean LMCI network and the group mean EMCI network.

Then we conducted feature extraction and classification experiments as described in **Section 3.3**. **Table 1** summarizes the classification performance, and **Table 2** lists the Student's *T* test *P*-values. Our results indicated both SVD and HO-SVD performed significantly better than raw features; moreover, our proposed HO-SVD has significant advantages in Accuracy, Specificity and Balanced Accuracy, compared to the standard SVD approach. It is well-known that FACT-derived brain networks have a fair amount of noise, due to the substantial proportion of false positive fibers generated. Our experimental results suggest that HO-SVD is quite effective in handling feature reduction for these noisy networks. This conclusion was further supported by conducting the same experiments on EMCI and LMCI's brain networks derived from another three variants of FACT - the 2nd-order Runge-Kutta method [11], and the interpolated streamline [14] and tensorline [13] methods.

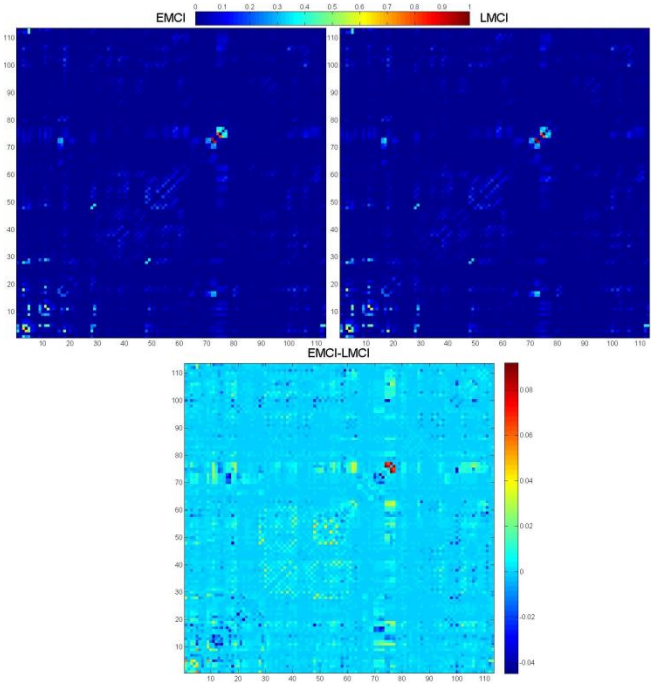

Figure 2. Mean brain network for EMCI, LMCI and the difference (EMCI-LMCI). Each cell of the network represents the connection strength (or connectivity) between each pair of ROIs; the ROIs are indexed from 1 to 113. (Please refer to [23] for ROI list details).

5. CONCLUSION

In this study, we proposed a novel framework to differentiate EMCI and LMCI using diffusion MRI derived structural networks, in conjunction with a sparse machine learning method. Experimental results indicated that our proposed framework performed better than a more traditional method (SVD or direct comparisons of matrix elements) in our network classification tests. Future studies will extend this framework to multi-task classification for all stages of Alzheimer's disease, as well as including data from other modalities (anatomical MRI, PIB-PET) that may assist with classification.

Table 1. Classification Performance for three feature extraction methods. Bal. ACC represents balanced accuracy.

	Raw	SVD	HO-SVD
Accuracy	0.59±0.12	0.59±0.08	0.65±0.07
Sensitivity	0.51±0.26	0.71±0.15	0.78±0.14
Specificity	0.61±0.16	0.58±0.09	0.64±0.08
AUC	0.55±0.13	0.68±0.10	0.70±0.10
Bal. ACC	0.56±0.10	0.64±0.09	0.71±0.07

Table 2. Student's *t* test *P*-value. "Bal. ACC" denotes the balanced accuracy. A Bonferroni correction was adopted here, to account for multiple comparisons. As there are five measures, the corrected *P* threshold in each column is 0.05/5=0.01. *P*-values less than 0.01 have been marked in *red*.

	SVD>Raw	HO-SVD>Raw	HO-SVD>SVD
Accuracy	0.53	0.03	6.46e-03
Sensitivity	2.68e-03	1.24e-04	0.08
Specificity	0.78	0.24	0.01
AUC	5.62e-04	9.91e-05	0.26
Bal. ACC	4.73e-03	2.61e-06	9.58e-03

ACKNOWLEDGMENTS

This work was supported in part by NIH *Big Data to Knowledge* (BD2K) Center of Excellence grant U54 EB020403, funded by a cross-NIH consortium including NIBIB and NCI.

REFERENCES

1. Petersen, R.C., et al., *Mild cognitive impairment: clinical characterization and outcome.* Arch Neurol, 1999. **56**(3): p. 303-8.
2. Grundman, M., et al., *Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials.* Arch Neurol, 2004. **61**(1): p. 59-66.
3. Aisen, P.S., et al., *Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans.* Alzheimers Dement, 2010. **6**(3): p. 239-46.
4. Leow, A., et al., *Impaired Inter-Hemispheric Integration in Bipolar Disorder Revealed with Brain Network Analyses.* Biological Psychiatry, 2012.
5. GadElkarim, J., et al., *A Framework for Quantifying Node-Level Community Structure Group Differences in Brain Connectivity Networks.* Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, 2012: p. 196-203.
6. Arienzo, D., et al., *Abnormal brain network organization in body dysmorphic disorder.* Neuropsychopharmacology, 2013. **38**(6): p. 1130-9.

7. Jahanshad, N., et al., *Disrupted brain networks in the aging HIV+ population.* Brain Connect, 2012.
8. Zhan, L., et al. *Brain network efficiency and topology depend on the fiber tracking method: 11 tractography algorithms compared in 536 subjects*. in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. 2013.
9. Zhan, L., et al., *Comparison of 9 Tractography Algorithms for Detecting Abnormal Structural Brain Networks in Alzheimer's Disease.* Submitted to PLOS ONE, 2014.
10. Zhan, L., et al., *Heritability of Brain network in 853 twins and sibilings.* Submitted to ISBI2015, 2014.
11. Basser, P.J., et al., *In vivo fiber tractography using DT-MRI data*. Magn Reson Med, 2000. **44**(4): p. 625-32.
12. Mori, S., et al., *Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging*. Ann Neurol, 1999. **45**(2): p. 265-9.
13. Lazar, M., et al., *White matter tractography using diffusion tensor deflection*. Hum Brain Mapp, 2003. **18**(4): p. 306-21.
14. Conturo, T.E., et al., *Tracking neuronal fiber pathways in the living human brain.* Proc Natl Acad Sci U S A, 1999. **96**(18): p. 10422-7.
15. Behrens, T.E., et al., *Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?* Neuroimage, 2007. **34**(1): p. 144-55.
16. Aganj, I., et al., *A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography.* Med Image Anal, 2011. **15**(4): p. 414-25.
17. Parker, G.J., H.A. Haroon, and C.A. Wheeler-Kingshott, *A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements.* J Magn Reson Imaging, 2003. **18**(2): p. 242-54.
18. De Lathauwer, L., B. De Moor, and J. Vandewalle, *A Multilinear Singular Value Decomposition.* SIAM Journal on Matrix Analysis and Applications, 2000. **21**(4): p. 1253-1278.
19. Mocks, J. and R. Verleger, *Principal component analysis of event-related potentials: a note on misallocation of variance.* Electroencephalogr Clin Neurophysiol, 1986. **65**(5): p. 393-8.
20. Tibshirani, R., *Regression shrinkage and selection via the lasso: a retrospective.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2011. **73**: p. 273–282.
21. Jahanshad, N., et al. *Diffusion tensor imaging in seven minutes: determining trade-offs between spatial and directional resolution*. 2010. IEEE.
22. Zhan, L., et al., *Angular versus spatial resolution trade-offs for diffusion imaging under time constraints.* Hum Brain Mapp, 2013. **34**(10): p. 2688-706.
23. Zhan, L., et al., *Magnetic resonance field strength effects on diffusion measures and brain connectivity networks.* Brain Connect, 2013. **3**(1): p. 72-86.
24. Desikan, R.S., et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.* Neuroimage, 2006. **31**(3): p. 968-80.
25. Liu, J., S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections.* Arizona State University, 2009. **http://www.public.asu.edu/~jye02/Software/SLEP**.