

# Nonlinear Metric Learning for Alzheimer’s Disease Diagnosis with Integration of Longitudinal Neuroimaging Features: Supplementary Material

Bibo Shi<sup>1</sup>  
bs354409@ohio.edu  
Yani Chen<sup>1</sup>  
yc147311@ohio.edu  
Kevin Hobbs<sup>2</sup>  
hobbsk@ohio.edu  
Charles D. Smith<sup>3</sup>  
csmith@mri.uky.edu  
Jundong Liu<sup>1</sup>  
liu@cs.ohio.edu

<sup>1</sup> School of Electrical Engineering and Computer Science  
Ohio University  
Athens OH, USA

<sup>2</sup> Department of Biological Sciences  
Ohio University  
Athens OH, USA

<sup>3</sup> Department of Neurology  
University of Kentucky  
Lexington KY, USA

## 1 Demographic and clinical information of the studied subjects

In our study, 338 subjects were selected for classifications: 94 of them are with AD, 121 with MCI and 123 are normal controls (NC). In this technical supplement, we first provide the details about the demographics and clinical evaluations, i.e., Mini Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) scores, of the studied subjects at their baseline visits, as shown in Table 1.

Diagnosis	Number	Gender (M/F)	Age (mean±sdv.) [min-max]	MMSE (mean±sdv.) [min-max]	CDR (mean±sdv.) [min-max]
AD	94	47/47	75.85 ± 7.2 [55 – 90]	23.21 ± 1.9 [20 – 26]	0.8 ± 0.25 [0.5 – 1]
MCI	121	69/52	75.73 ± 7.8 [55 – 90]	26.57 ± 1.7 [23 – 30]	0.5 ± 0 [0.5 – 0.5]
NC	123	65/58	76.08 ± 5.2 [62 – 90]	29.15 ± 0.9 [26 – 30]	0 ± 0 [0 – 0]

Table 1: Demographic and clinical information of the studied subjects at the baseline.

© 2015. The copyright of this document resides with its authors.  
It may be distributed unchanged freely in print or electronic forms.

## 2 Selected patches for “GM only” and “DM only” features

In this section, we provide additional visualizations of the selected patches for “GM only” and “DM only” features, as shown in Fig 1 and Fig 2, respectively.

Similar to the selected patches for “Joint GM & DM” feature, the most discriminative regions detected by “GM only” and “DM only” features both include hippocampus, parahippocampal gyrus, entorhinal cortex, and amygdala, which are consistent with the findings reported in [10, 8, 10]. Differently from “Joint GM & DM”, each single type of feature captures much more patches; also, the “DM only” feature captures more patches in the outer cortical area than “GM only” feature.

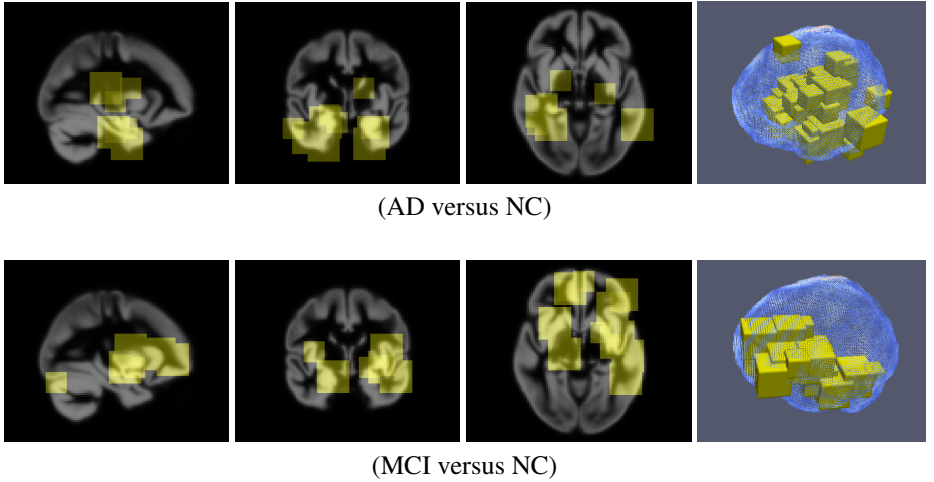


Figure 1: The selected patches in “GM only” feature for AD vs. NC and MCI vs. NC classifications. The columns from left to right are sagittal, coronal, axial and 3D views.

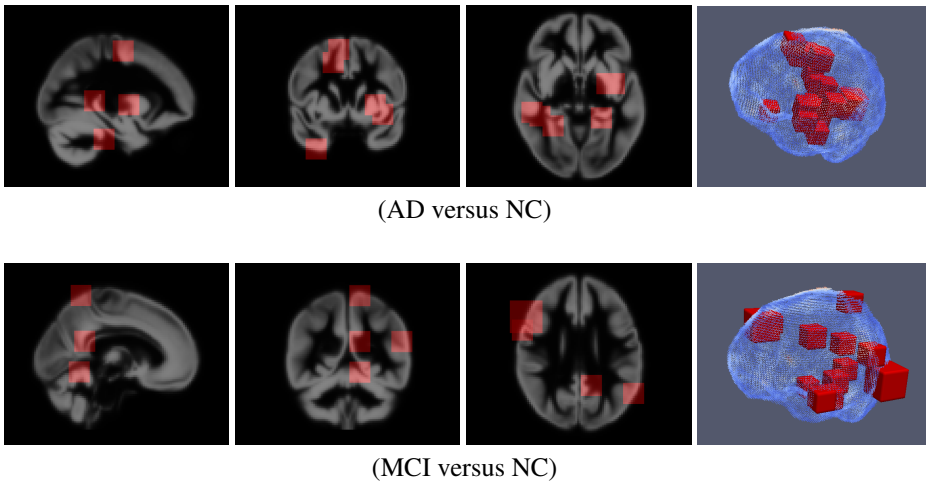


Figure 2: The selected patches in “DM only” feature for AD vs. NC and MCI vs. NC classifications. The columns from left to right are sagittal, coronal, axial and 3D views.

## 3 Comparisons of ML-TPS with other ML methods on UCI data

In this section, we present two additional sets of experiments to evaluate the classification performance of the proposed nonlinear ML-TPS on seven widely used datasets from UCI machine learning repository. The leftmost column of Table 2 summarizes the details of the datasets. All datasets have been preprocessed through normalization.

### 3.1 NN based comparisons on UCI datasets

The vast majority of the existing distance metric learning solutions [2, 3, 4, 5, 6, 8] were designed and can be applied to improve metric based classification methods, especially the Nearest Neighbor (NN) classifiers. Thus, in the first set of experiments, we choose  $k$ NN method ( $k = 1$ ) as the baseline classifier, and compare the improvements made by ML-TPS against five state-of-the-art metric learning methods: Large Margin Nearest Neighbor classification (LMNN) method [4], Information-Theoretic Metric Learning (ITML) method [2], Neighborhood Components Analysis (NCA) method [5], multi metric LMNN (mm-LMNN) [8] and Parametric Local Metric Learning (PLML) method [6]. The hyper-parameters of NCA, ITML, LMNN and mm-LMNN are set by following [2, 3, 4, 8] respectively. PLML has a number of hyper-parameters, so we follow the suggestion of [6]: use a 3-fold CV to select  $\alpha_2$  from  $\{0.01 \sim 1000\}$ , and set the other hyper-parameters by its default. In our ML-TPS model, there are two hyper-parameters: the number of anchor points  $p$  and the weighting factor  $\lambda$ . For  $p$ , we empirically set it to 30% of the training samples; for  $\lambda$ , we select it through CV from  $\{5^{-5} \sim 5^{25}\}$ .

Datasets [#Inst./#Feat./#Class]	kNN	LMNN	ITML	NCA	PLML	mm-LMNN	ML-TPS
Iris [150/4/3]	95.70 $\pm$ 2.31 (4.0)	95.06 $\pm$ 2.62 (3.5)	95.22 $\pm$ 2.56 (3.5)	94.68 $\pm$ 2.35 (3.0)	84.22 $\pm$ 4.54 (0.0)	93.60 $\pm$ 2.68 (1.0)	96.49 $\pm$ 2.32 +++++ (6.0)
Wine [178/13/3]	95.21 $\pm$ 2.04 (0.5)	97.25 $\pm$ 1.80 (5.0)	96.90 $\pm$ 2.31 (4.0)	96.65 $\pm$ 2.27 (3.5)	96.61 $\pm$ 2.10 (3.0)	95.16 $\pm$ 2.53 (0.5)	97.18 $\pm$ 2.05 +++++ (4.5)
Breast [683/10/2]	95.35 $\pm$ 1.34 (1.0)	95.66 $\pm$ 1.39 (2.5)	95.76 $\pm$ 1.30 (2.5)	95.57 $\pm$ 1.13 (1.5)	96.18 $\pm$ 0.98 (5.0)	96.13 $\pm$ 1.13 (5.0)	95.97 $\pm$ 1.04 +++++ (4.0)
Diabetes [768/8/2]	70.58 $\pm$ 2.26 (4.5)	70.54 $\pm$ 2.52 (4.5)	68.81 $\pm$ 2.65 (1.0)	68.53 $\pm$ 2.71 (1.0)	69.04 $\pm$ 2.30 (1.5)	69.68 $\pm$ 2.53 (2.5)	71.54 $\pm$ 2.21 +++++ (6.0)
Liver [345/6/2]	61.20 $\pm$ 3.96 (2.0)	60.79 $\pm$ 3.54 (2.0)	60.07 $\pm$ 4.92 (1.5)	62.63 $\pm$ 4.15 (4.0)	64.74 $\pm$ 3.99 (5.5)	59.48 $\pm$ 3.93 (0.5)	64.00 $\pm$ 4.36 +++++ (5.5)
Sonar [208/60/2]	84.73 $\pm$ 3.45 (3.0)	84.12 $\pm$ 4.13 (2.0)	82.14 $\pm$ 5.94 (0.0)	85.46 $\pm$ 3.51 (3.5)	87.42 $\pm$ 4.70 (6.0)	84.68 $\pm$ 3.94 (3.0)	85.35 $\pm$ 3.82 +++++ (3.5)
Ionosphere [846/18/4]	85.83 $\pm$ 2.62 (0.0)	88.40 $\pm$ 2.54 (3.0)	87.45 $\pm$ 3.07 (1.0)	88.33 $\pm$ 2.77 (3.0)	91.03 $\pm$ 2.23 (5.5)	91.68 $\pm$ 1.13 (5.5)	88.39 $\pm$ 2.37 +++++ (3.0)
Total Score	15.0	22.0	13.5	19.5	26.5	18.0	32.5

Table 2: Mean and standard deviation of classification accuracy results on seven UCI datasets. The superscripts  $+-$  in ML-TPS column indicate a significant win, loss or no difference respectively based on the pairwise Student’s  $t$ -test with the other six methods. The number in the parenthesis denotes the score of the respective method for the given dataset.

To better compare the classification performance, we run the experiment 100 times with different random 3-fold splits of each dataset, two for training and one for testing. Furthermore, we conduct a pairwise Student’s  $t$ -test with a  $p$ -value 0.05 among the seven methods for each dataset. Then, a simple ranking scheme [9] is used to evaluate the relative performance of those methods: a method A will be assigned 1 point if it has a statistically

significantly better accuracy than another method B, 0.5 points if it has no significant difference, and 0 point if it is significantly worse than B. The experimental results by averaging over the 100 runs along with the ranking scores are reported in Table 2.

From Table 2, we can see that ML-TPS outperforms all the other six methods in a statistical significant manner, with a total score of 32.5 points. Out of the total 42 pair-wise comparisons, ML-TPS has 26 statistical wins in total. Furthermore, ML-TPS has significantly improved the performance of the baseline  $k$ NN classifier on six out of all seven datasets, and performed equally well on the seventh (“Sonar”).

### 3.2 SVM based comparisons on UCI datasets

Recently, Xu, Weinberger and Chapelle [9] pointed out that metric learning methods can be used as the preprocessing step to transform the feature space for, or combined with SVMs to improve SVMs’ performance. In the second set of experiments, we choose the Gaussian kernel SVM ( $k$ SVM) as the baseline method, compare the improvements made by ML-TPS against LMNN [9], ITML [9], and NCA[9], as the preprocessing feature transformation step for  $k$ SVM. Note that the multi-metric learning methods are not easily generalized to SVM, so we didn’t consider the two multi-metric learning methods PLML and mm-LMNN here.

For  $k$ SVM, we tune the two hyper-parameters  $C$  and  $\sigma$  via 3-fold inner cross validation (CV) respectively from  $\{2^{-15} \sim 2^{15}\}$ . The hyper-parameters of NCA, ITML, LMNN are set in the same way as in the NN based experiments. We also adopt the same experimental setting and statistical ranking schema as in the NN based classification, and report the results in Table 3.

Datasets [#Inst./#Feat./#Class]	$k$ SVM	LMNN	ITML	NCA	ML-TPS
Iris [150/4/3]	96.02 $\pm$ 2.29 (1.0)	96.69 $\pm$ 2.32 (3.5)	95.75 $\pm$ 2.82 (1.0)	96.00 $\pm$ 2.31 (1.0)	96.80 $\pm$ 2.48 ++++ (3.5)
Wine [178/13/3]	97.54 $\pm$ 1.58 (2.0)	97.63 $\pm$ 1.86 (2.5)	97.68 $\pm$ 1.43 (2.5)	97.04 $\pm$ 2.11 (0.5)	97.85 $\pm$ 1.89 ====+ (2.5)
Breast [683/10/2]	96.67 $\pm$ 1.00 (1.5)	97.00 $\pm$ 1.01 (3.5)	96.48 $\pm$ 1.10 (1.5)	96.14 $\pm$ 1.02 (0.0)	97.25 $\pm$ 0.99 ++++ (3.5)
Diabetes [768/8/2]	77.17 $\pm$ 2.06 (2.5)	76.92 $\pm$ 2.08 (2.5)	76.79 $\pm$ 2.30 (2.5)	76.00 $\pm$ 2.21 (0.0)	77.14 $\pm$ 2.06 ====+ (2.5)
Liver [345/6/2]	72.48 $\pm$ 3.30 (3.0)	71.95 $\pm$ 3.37 (3.0)	70.84 $\pm$ 3.26 (1.5)	68.96 $\pm$ 4.30 (0.0)	71.56 $\pm$ 3.56 ====+ (2.5)
Sonar [208/60/2]	86.07 $\pm$ 3.68 (3.0)	82.68 $\pm$ 4.74 (1.5)	80.56 $\pm$ 7.33 (0.0)	83.46 $\pm$ 4.23 (1.5)	88.28 $\pm$ 3.91 ++++ (4.0)
Ionosphere [846/18/4]	94.44 $\pm$ 1.83 (2.0)	92.23 $\pm$ 2.81 (0.0)	97.37 $\pm$ 3.10 (4.0)	93.79 $\pm$ 2.01 (1.0)	95.25 $\pm$ 1.72 ++++ (3.0)
Total Score	15.0	16.5	13.0	4.0	21.5

Table 3: Mean and standard deviation of SVMs based classification accuracy results on seven UCI datasets. The settings and notations of the comparison scores are identical to those in Table 1.

From Table 3, we can see that ML-TPS outperforms all the other four methods in a statistical significant manner, with a total score of 21.5 points. Out of the total 28 pair-wise comparisons, ML-TPS has 16 statistical wins in total. Furthermore, it is obvious that adding the proposed ML-TPS as the preprocessing step has significantly improved the performance of  $k$ SVM, with better classification accuracies on four out of all seven datasets (“Iris”, “Breast”,

“Sonar”, and “Ionosphere”), and doing comparably well on the other three datasets. Considering the already state-of-the-art performance of  $k$ SVM (ranking the 3rd among all the five methods), this improvement made by ML-TPS is very significant, while the other three metric learning methods (LMNN, ITML, NCA) even degrade the performance of  $k$ SVM on some datasets (“Liver”, “Sonar”, or “Ionosphere”).

## References

- [1] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehericy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer’s Disease Neuroimaging Initiative, et al. Automatic classification of patients with alzheimer’s disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, 56(2):766–781, 2011.
- [2] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007.
- [3] Yi Hong, Quannan Li, Jiayan Jiang, and Zhuowen Tu. Learning a mixture of sparse distance metrics for classification and dimensionality reduction. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 906–913. IEEE, 2011.
- [4] Sam Roweis Jacob Goldberger and Ruslan Salakhutdinov Geoff Hinton. Neighbourhood components analysis. *NIPS’04*, 2004.
- [5] Manhua Liu, Daoqiang Zhang, and Dinggang Shen. Hierarchical fusion of features and classifier decisions for alzheimer’s disease diagnosis. *Human brain mapping*, 35(4):1305–1319, 2014.
- [6] Jun Wang, Alexandros Kalousis, and Adam Woznica. Parametric local metric learning for nearest neighbor classification. In *Advances in Neural Information Processing Systems*, pages 1601–1609, 2012.
- [7] Kilian Weinberger, John Blitzer, and Lawrence Saul. Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems*, 18:1473, 2006.
- [8] Kilian Q Weinberger and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009.
- [9] Zhixiang Xu, Kilian Q Weinberger, and Olivier Chapelle. Distance metric learning for kernel machines. *arXiv preprint arXiv:1208.3422*, 2012.
- [10] Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Multimodal classification of alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011.