Boosting power for clinical trials using classifiers based on multiple biomarkers

Omid Kohannim[a], Xue Hua PhD[a], Derrek P. Hibar BS[a], Suh Lee BS[a], Yi-Yu Chou MS[a], Arthur W. Toga PhD[a], Clifford R. Jack Jr MD[b], Michael W. Weiner MD[c,d,e], Paul M. Thompson PhD[a] and the Alzheimer's Disease Neuroimaging Initiative*

[a]Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, Los Angeles, CA, USA

[b]Dept. of Radiology, Mayo Clinic, Rochester, MN, USA

[c]Dept. of Radiology and Biomedical Imaging, UCSF, San Francisco, CA, USA

[d]Dept. of Medicine, UCSF, San Francisco, CA, USA

[e]Dept. of Psychiatry, UCSF, San Francisco, CA, USA

Introduction:

Alzheimer's Disease (AD) is the most common form of dementia, affecting 5.3 million people in the United States. Multiple imaging and other biomarkers have been used for quantifying disease progression, measuring various aspects of AD pathology. We implemented support vector machines (SVMs) as a machine learning tool to classify patients as having AD and mild cognitive impairment (MCI) – an intermediate condition between healthy aging and AD – and to predict future cognitive decline based on combinations of biomarkers including imaging. We also used our classifiers to predict those most likely to decline, to seeks out a sub-sample of subjects who might be better candidates for demonstrating therapeutic effects in a clinical trial. We hypothesized that using combinations of biomarkers would offer complementary information to classify

patients into the correct diagnostic categories and predict cognitive decline, thereby providing a new way to boost the power of clinical trials.

Methods:

Demographic and biomarker data were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database (http://www.loni.ucla.edu/ADNI/Data/). These included age, sex, body mass index (BMI), ApoE, baseline CSF biomarker measures (for $A\beta_{42}$, t-tau and p-tau), a baseline, temporal PET-FDG summary (Landau et al.), and sum of boxes CDR clinical scores. The baseline MRI measures included numerical summaries from the hippocampus, lateral ventricles and a measure of atrophy in the temporal lobes. The hippocampal summaries were volumes generated from an automatic segmentation method that we developed based on machine learning (Morra et al., 2009). The ventricular summaries were volumes acquired from a semi-automated, multi-atlas segmentation technique that we developed (Chou et al., 2009). The temporal lobe summaries were obtained from an anatomically defined region-of-interest (ROI) on 3D maps generated with tensor-based morphometry (Hua et al., 2008a, Hua et al., 2008b). Sample sizes for analyses using different predictors were slightly different, as not all measures could be collected from all ADNI subjects. Subjects for each analysis were randomly divided into training and testing sets for the SVM experiments.

SVMs are a type of machine learning method that can be used to classify datasets into different groups, based on combining multiple features, or measures, available in each subject. Through SVM, a number of observations about a subject may be assembled into a vector, with as many components as there are measures. Then a mathematical function

is estimated (or "learned") that best combines these features to give an output that indicates which group the individual belongs to.

Power analysis was performed as in our prior studies (Hua et al., 2009). We estimated the number of subjects needed, in each arm of a hypothetical clinical trial, to detect a 25% reduction in the average atrophy rate, with 80% power (referred to as n80).

Results:

In our first study with 635 subjects, we performed SVM training using MRI summaries, ApoE, age, sex and BMI, ranked the features based on SVM weights, and found the top features that gave the highest classification accuracy. For AD vs. control, the top 4 features yielded the highest accuracy of 82.2%, and an area under the ROC curve (AUC) of 0.945. For MCI vs. control, the best features comprised the top 3, with 70.9% accuracy and an AUC of 0.860. The rank orders and best biomarker sets are displayed in **Table 1**. **Figure 1** shows the ROC curves.

Next, we considered two subsets of the subjects above by adding CSF and PET-FDG separately, and obtained new rank orders, shown in **Table 2.** CSF t-tau and $a\beta_{42}$ as well as PET-FDG were among the best set of biomarkers for both AD and MCI classification. Including CSF or PET-FDG in the top list of features improved classification accuracy, implying that PET-FDG and CSF complemented MRI, ApoE and age. The AUCs, however, did not change significantly, perhaps due to the small size of the testing sets.

Lastly, we computed minimum sample size estimates (n80) for the top *k* percent of subjects classified as most likely to have AD with our classifier, using the top 4 features highlighted in Table 1. When *k* is less than 33%, the power estimates for AD subjects are

improved compared to the sample size of 48 subjects reported by Hua et al. (2009) **(Figure 2a)**. To obtain similar estimates for MCI, we changed the output of our SVM algorithm to the 1-year change in sobCDR, instead of a binary output. With PET-FDG, MRI ventricular and temporal summaries, and ApoE identified as the best set of predictors, we trained a model that predicted sobCDR change in a testing set. We ranked the test subjects in order of predicted cognitive decline and computed n80 estimates for the top *k*% percent of MCI subjects (**Figure 2b**). The n80 estimates were lower than the 88 individuals reported by Hua et al. (2009) as the minimal sample size for MCI.

Conclusions:

We determined combinations of regional MRI numerical summaries with demographic variables and ApoE that best classified AD or MCI vs. control. Combining MRI, ApoE and age was very useful for AD and MCI classification, but more so for AD, in agreement with what is known about the use of structural MRI in AD (Frisoni et al., 2010). When compared to PET-FDG and CSF, MRI measures contributed most to AD classification, whereas PET-FDG and CSF biomarkers, particularly $a\beta_{42}$ were more helpful for MCI classification. This is also consistent with recent hypotheses on the temporal sequence of the dynamic trajectories of biomarkers for AD (Jack et al., 2010, Petersen et al., 2010). Using an AD classifier and an MCI cognitive decline predictor based on combinations of biomarkers, we reduced clinical trial sample estimates to fewer than 40 AD and MCI subjects to detect a 25% slowing in temporal lobe atrophy rates with 80% power - a substantial boosting of power relative to standard imaging measures.

Figures and Tables:

Table 1 – Rank order list for MRI, ApoE, Age, Sex and BMI in AD and MCI classification.

	Biomarker	
Rank	AD vs. CN	MCI vs. CN
1	MRI Hip[a]	MRI Hip
2	ApoE	ApoE
3	Age	Age
4	MRI Vent[b]	MRI Vent
5	MRI Temp[c]	MRI Temp
6	BMI	BMI
7	Sex	Sex

[a]Hippocampal volume summary
[b]Ventricular volume summary
[c]Temporal lobe summary from tensor-based morphometry (TBM)
Groups of biomarkers yielding the highest leave-one-out accuracy are highlighted.

Figure 1. ROC curves for AD and MCI classification.

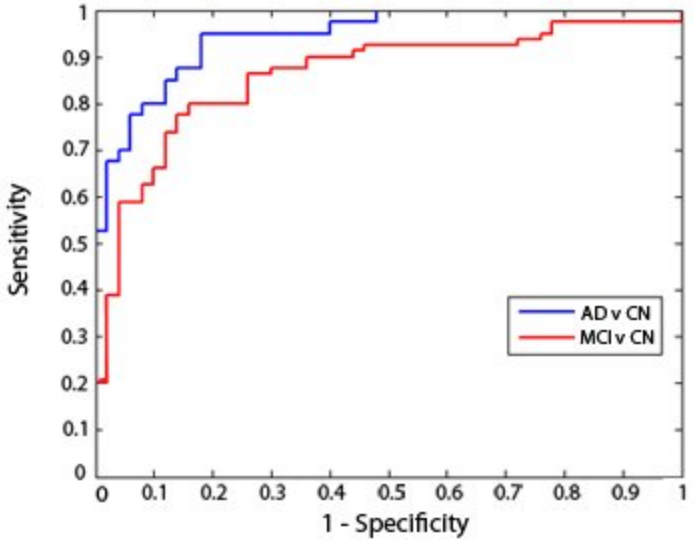
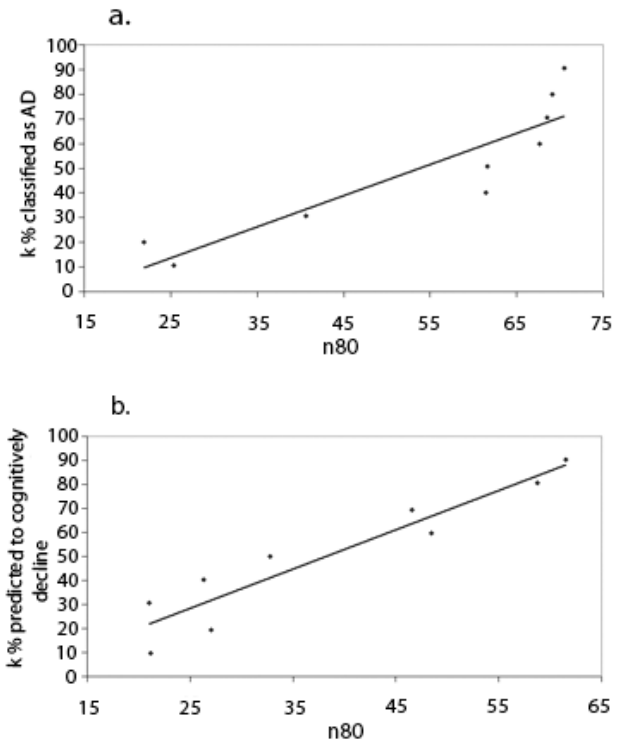


Table 2 – Rank order list for MRI, ApoE, Age, Sex, BMI and either (a) CSF or (b) PET-FDG, for AD and MCI classification.

	Biomarker			
	a. MRI + CSF		b. MRI + PET-FDG	
Rank	AD vs. CN	MCI vs. CN	AD vs. CN	MCI vs. CN
1	MRI Hip[a]	MRI Hip	ApoE	ApoE
2	CSF t-tau	CSF $a\beta_{42}$	PET-FDG	MRI Hip
3	CSF $a\beta_{42}$	Age	MRI Hip	PET-FDG
4	ApoE	ApoE	MRI Vent	Age
5	MRI Vent[b]	CSF t-tau	Age	MRI Temp
6	Age	Gender	MRI Temp	Sex
7	MRI Temp[c]	MRI Temp	BMI	MRI Vent
8	BMI	BMI	Sex	BMI
9	CSF p-tau	CSF p-tau		
10	Sex	MRI Vent		

[a]Hippocampal volume summary
[b]Ventricular volume summary
[c]Temporal lobe summary from tensor-based morphometry (TBM)
Sets of biomarkers yielding the highest leave-one-out accuracy are highlighted.

Figure 2. n80 estimates for (a) AD, and (b) MCI



References:

Chou, Y.-Y., Leporé, N., et al. (2008), "Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease", NeuroImage, vol. 40, pp. 615-630.

Frisoni, G. B., Fox, N. C., et al. (2010), "The clinical use of structural MRI in Alzheimer disease", Nature Reviews | Neurology, vol. 6, pp. 1-11.

Hua, X., Lee, S., et al. (2009), "Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: An ADNI study of 515 subjects", NeuroImage, vol. 48, pp. 668-681.

Hua, X., Leow, A. D., et al. (2008b), "Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects", NeuroImage, vol. 43, pp. 458-469.

Hua, X., Leow, A. D. (2008a), "3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry", NeuroImage, vol 41, pp. 19-34.

Jack, C. R., Knopman, D.S.K., et al. (2010), "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade", Lancet Neurology, vol. 9, pp. 119-128.

Morra, J. H., Tu, Z., et al. (2009), "Automated 3D Mapping of Hippocampal Atrophy and Its Clinical Correlates in 400 Subjects with Alzheimer's Disease, Mild Cognitive Impairment, and Elderly Controls", Human Brain Mapping, vol. 30, pp. 2766-2788.

Petersen, R. C. (2010), "Alzheimer's disease: progress in prediction", Lancet Neurology, vol. 9, pp. 4-5.