Mapping Hippocampal Degeneration in 400 Subjects with a Novel Automated Segmentation Approach

Jonathan Morra[1], Zhuowen Tu[1], Liana Apostolova [1,2], Amity Green [1,2], Christina Avedissian[1], Sarah Madsen[1], Neelroop Parikshak[1], Xue Hua[1], Arthur Toga[1], Clifford Jack[3], Norbert Schuff[4], Michael Weiner [4,5], Paul Thompson [1]

[1]Laboratory of Neuro Imaging, UCLA, Los Angeles, CA, United States/[2]Dept. of Neurology, UCLA, Los Angeles, CA, United States /[3]Mayo Clinic College of Medicine, Rochester, MN, United States/[4]Dept, of Radiology, UCSF, San Francisco, CA, United States/ [5]Dept. of Medicine and Psychiatry, UCSF, San Francisco, CA, United States

Introduction: We automatically segmented the hippocampus in 400 brain MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI [1]), using a novel segmentation approach that combines AdaBoost with a novel model, called the Auto Context Model (ACM). Our goal was to create statistical maps correlating hippocampal morphology with different covariates of interest, including diagnosis (normal, mild cognitive impairment (MCI), Alzheimer's Disease (AD)), mini-mental state examination (MMSE) scores, and global and "sum-of-boxes" clinical dementia ratings (gCDR and sobCDR).

Methods: As part of a multi-site longitudinal neuroimaging study [1] we analyzed 400 3D T1-weighted brain MRI scans from 100 normal elderly, 200 MCI, and 100 AD subjects (age: 75.8+/-6.6SD). Clinical measures included MMSE scores (scores <24/30 typically indicate dementia), gCDR scores (0, 0.5, 1, 2, and 3, indicate no dementia, very mild, mild, moderate, or severe dementia), and sobCDR scores (0=no dementia; 18=very severe dementia). In the training phase, our algorithm was provided with 21 hand-labeled segmentations, and a classification rule for hippocampal versus non-hippocampal regions was learned using a modified AdaBoost method, based on a pool of ~13,000 features, including intensity, combinations of x, y, and z positions, curvatures, gradients, tissue classification maps of gray matter, white matter, and CSF, mean filters, standard deviation filters, and Haar filters of size 1x1x1 to 7x7x7. We linearly registered all brains to a standard template to devise a basic shape prior to capture global hippocampal shape, defined as the pointwise summation of all the training masks. We also included curvature, gradient, mean, standard deviation, and Haar filters of the shape prior as features. Our extension of AdaBoost, the Auto Context Model, is described in **Figure 1**. In ACM, the posterior distribution for the Adaboost classification rule was iteratively updated by feeding the current estimate of the posterior distribution back in as a new input feature for AdaBoost to use as a weak learner. After about four iterations of ACM, the posterior distribution converges, yielding a final segmentation in less than 1 minute on a desktop computer.

Based on these hippocampal segmentations, we correlated hippocampal shape with different clinical covariates using surface-based statistical maps. 3D parametric surface models were constructed from each binary segmentation, and geometrically averaged across subjects within each diagnostic group. We applied a radial distance atrophy mapping approach to examine 3D structural and volumetric differences between groups. We employed a surface averaging approach to establish pointwise correspondence, and created significance maps (*p*-maps) for the effects of diagnosis and clinical measures.

Results: **Figure 2** shows the *p*-maps for (1) each pairwise diagnostic comparison, and (2) correlating atrophy with MMSE, gCDR, and sobCDR scores. The level of atrophy was strongly associated with MMSE and CDR scores, and with diagnosis (with greatest effects for the AD v. normal comparison).

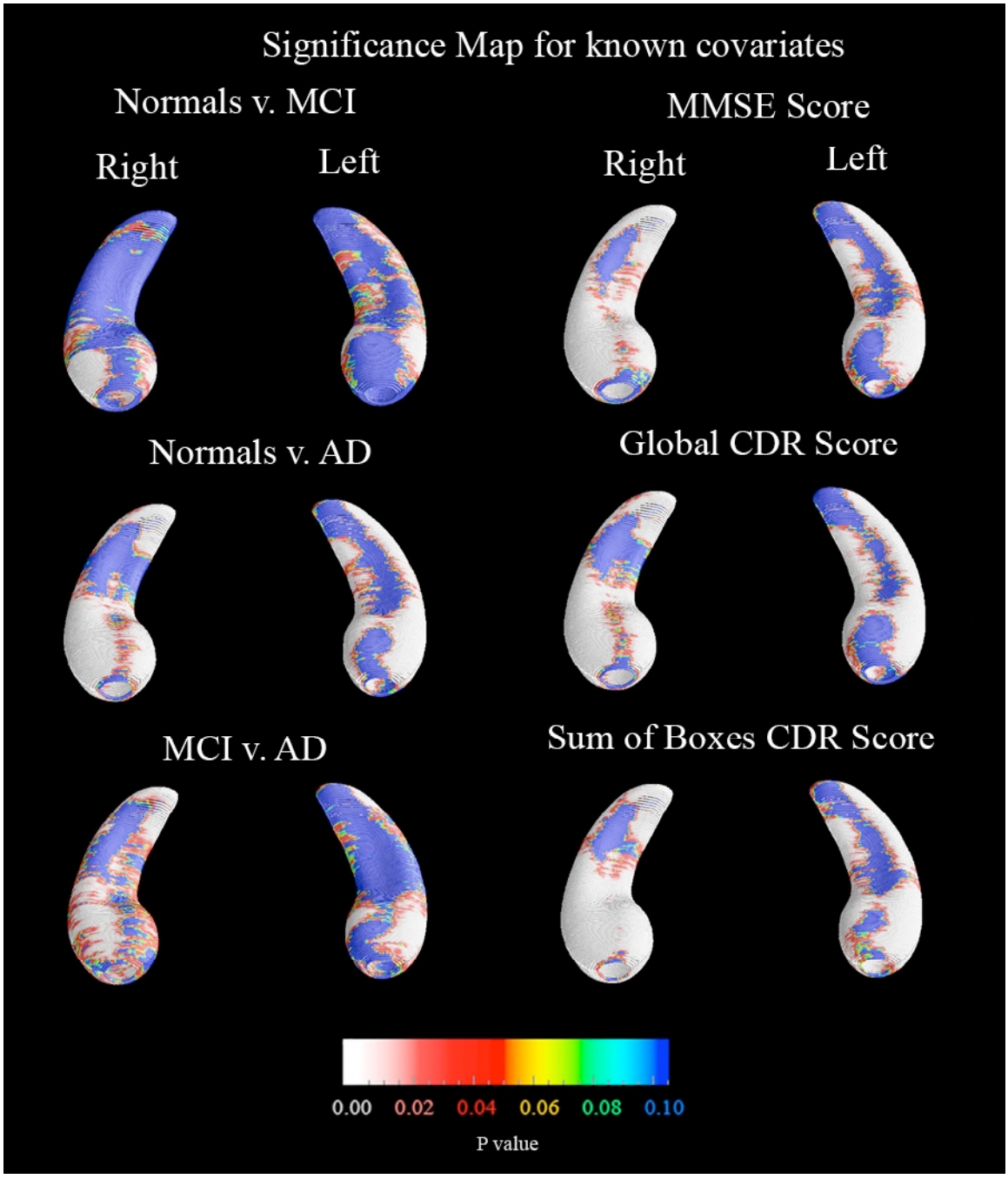
Conclusions: AdaBoost, extended here using ACM, selects features based on a training set of expert segmentations, and can accurately detect statistical linkages between diagnosis and clinical scores and radial atrophy. This is one of the first studies to quantify neurodegeneration automatically in such a large cohort, using surface-based maps rather than simple volumetric measures.

References: [1]. www.loni.ucla.edu/ADNI

Given: N labeled training examples (x_i, y_i) with $y_i \in \{-1, +1\}$ and $x_i \in \mathbb{R}$, a set of J weak learners $h_j \in H$, and a prior distribution $P_1(i)$
For $s = 1, \ldots, S$:

- Create an initial uniform distribution of weights $D_1(i)$ over the examples
- For $t = 1, \ldots, T$:
  - $\epsilon_j = \sum_{i=1}^{N} D_t(i)\mathbf{1}(y_i \neq h_j(x_i))$
  - $[h_t, \epsilon_t] = \arg \min_{h_j \in H} \epsilon_j$
  - Set $\alpha_t = \frac{1}{2} \log\left((1 - \epsilon_t)/\epsilon_t\right)$
  - Set $D_{t+1}(i) = D_t(i) exp(-\alpha_t y_i h_t(x_i))/Z_t$
    $Z_t = 2\sqrt{\epsilon_t(1 - \epsilon_t)}$, a normalization factor
- Calculate $P_s(x) = 1/\left(exp(-f(x)) + 1\right)$
  $f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$
- Update those weak learners based on $P_s(i)$

Fig. 1. The AdaBoost algorithm inside of the Auto Context Model. $\mathbf{1}$ is an indicator function. Weak learners h_t are assumed to output $\{-1, +1\}$. In our context, x_i can be any feature, hence it is a member of $\mathbb{R}$.


Significance Map for known covariates
Normals v. MCI
Right
Left
MMSE Score
Right
Left
Normals v. AD
Global CDR Score
MCI v. AD
Sum of Boxes CDR Score
0.00
0.02
0.04
0.06
0.08
0.10
P value