

DISCOVERY OF "BIOMARKERS" FOR ALZHEIMER'S DISEASE PREDICTION FROM STRUCTURAL MR IMAGES

Y. Liu^{1,2,3}, L. A. Teverovskiy², O. L. Lopez³, H. Aizenstein³, C. C. Meltzer^{3,4}, J. T. Becker³

¹ Penn State University, ² Carnegie Mellon University, ³ University of Pittsburgh, ⁴ Emory University

ABSTRACT

We propose a computational framework for learning predictive image features as "biomarkers" for Alzheimer's Disease discrimination using high-resolution Magnetic Resonance (MR) brain images. We focus on the exploration of a very large (> 500 million) feature space derived extensively from the deformation and tensor fields. In such a huge space, our computational tool supports an automatic search for discriminative feature subspaces and the corresponding anatomical regions in human brains, which can be used to discriminate previously unseen, individual structural MR images from Alzheimer's Disease (AD) and normal control (CTL) subjects. Our aggressive leave-ten-out cross-validations on 40 subjects demonstrate higher than 90% sensitivity and specificity. In addition, we demonstrate intriguing anatomical locations as automatically discovered "biomarkers" and the spatial distributions of 20 Mild cognitive impairment (MCI) subjects in the discriminative feature space automatically learned for AD and CTL separations. Our results illustrate a truly complementary effort of human and computers for early diagnosis of AD from MR images.

Index Terms— Image classification, feature extraction, MRI, brain asymmetry,

1. INTRODUCTION

As a disease of aging, the financial and social burdens of Alzheimer's disease (AD) are compounded by the continuing increases in the average life span [1]. The prevalence of AD is estimated to continue to climb at a rapid rate, with an expected quadrupling of cases in the United States over the next 50 years [1]. Thus, the need for developing early diagnostic markers to complement new therapeutic approaches is more acute than ever before. These data have created an imperative for critically evaluating non-invasive tools with early diagnostic capability.

Structural magnetic resonance (MR) images of the brain have an important advantage over other imaging modalities as potential biomarkers in that they are non-invasive and provide

detailed information about gray and white matter parenchyma of the brain, and cerebrospinal fluid (CSF)-filled spaces.

Researchers have used both deformation and tensor fields to indicate the statistical differences in brain regions between normal controls and patients, for example [2, 3, 4, 5, 6]. Our work differs from previous work in that (1) we go beyond simply using the deformation and tensor fields as they are. Instead, we extract a substantial amount (> 500 millions) of content rich local and global "image features" from such 3D vector and scalar fields. For example, we compute the quantified brain asymmetry in multi-scale spaces. (2) we have developed a set of simple but effective computational tools that train the computers to automatically search for the discriminative subspaces of low dimensions (less than 5) from the original 500 million plus dimensions (significantly different from the state of the art non-supervised manifold learning for classification); (3) we go beyond finding significant p-values between controls and patients in certain brain regions. Instead, we tested the effectiveness of the automatically learned discriminative features, "MR image biomarkers", using unseen, one-time individual MR images with one of the most stringent cross validation strategies used for this type of study: leave-ten-out from 40 random splits of the data instead of the commonly used leave-one-out tests; (4) we demonstrate the precise anatomical locations and types of features of the "MR image biomarkers" that either echo the findings in the literature or expose novel, intriguing new sites leading to further exploration of potentially new biological interpretations and discoveries.

2. OUR METHOD

There are four major steps in our general framework for computer training, which is summarized in Figure 1.

Image alignment: Input MR images are processed through a predefined pipeline: intensity normalization, midsagittal plane (MSP) extraction [7] followed by affine deformation, and finally each MR image is fully deformably registered to the colin27 [8] digital atlas, using a finite element mesh based algorithm [9] followed by the Demons algorithm [10], both implemented in ITK [11].

Feature extraction: Instead of using the deformation field D and the tensor fields from the registration directly, we project

The work is supported by NIH P50-AG05133, MH064625, MH01077, AG05133, DA015900-01, and PA Health Department grant 4100027294

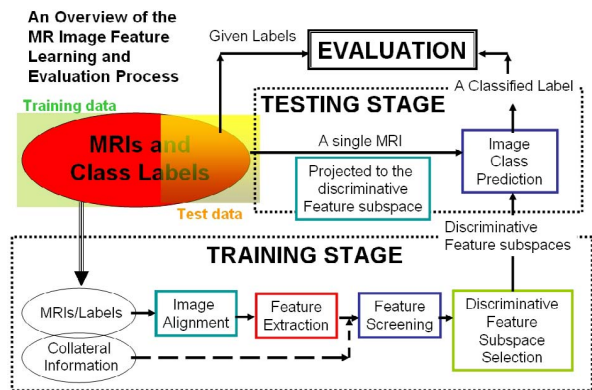


Fig. 1. An overview of our computational framework.

the 3D-vector fields into seven scalar fields: its X,Y,Z components, θ , ϕ angles, the vector length, and the determinant of the Jacobian matrix of D . Furthermore, we use four levels of image pyramids. and four levels of voxel neighborhoods and within each neighborhood compute six types of statistical measures including: minimum, maximum, mean, variance, 3rd and 4th moments, plus all their corresponding asymmetry w.r.t. the MSP. Therefore, the total number of features reaches 584,617,824.

Feature screening: We use an *augmented variance ratio* (AVR) as $AVR(F) = \frac{Var(S_F)}{\frac{1}{C} \sum_{i=1..C} \frac{Var_i(S_F)}{\min_{i \neq j} (|mean_i(S_F) - mean_j(S_F)|)}}$ where

$Var(S_F)$ is the cross-class variance of feature F , $Var_i(S_F)$ and $mean_i(S_F)$ are the within-class variance and mean of feature F for class i out of C distinct classes. Our experimental results demonstrate that AVR turns out to be an effective measure for our purposes in this experiment. Figure 3 demonstrates a pair of features is their relative AVR spaces where one of them is a good biomarker to discriminate AD from normal subjects.

Subspace biomarker learning and cross-validation: In addition to feature screening to rule out irrelevant feature dimensions, we have studied extensively the combination of different feature dimensions to form the most discriminative feature subspaces. Using different types of classifiers from K-nearest neighbor, decision tree, linear discriminative analysis (LDA) to SVM and boosting, we found that the simplest ones are the most effective (Section 3). We use the more stringent and statistically convincing 40-fold leave-ten-out cross validation instead of LOO.

3. EXPERIMENTAL RESULTS

3.1. Data

A set of 66 subjects are selected by experts from the Alzheimer's Disease Research Center. There are 20, 26 and 20 subjects

in normal controls (CTL), mild cognitive impairment (MCI) and diagnosed probable Alzheimer's Disease (AD) category respectively, matched on age, education and sex.

The MR image data were obtained on a 1.5 Tesla GE Signa scanner (GE Medical Systems, Milwaukee, MN) and acquired in three dimensions to obtain 120 thin, contiguous images throughout the entire brain. The contrast was designed to maximize the gray-white matter and CSF differences (TR=25, TE=5, slices > 1.5mm, 0 gap, 40E angle, FOV=24x18).

3.2. Results

During leave-ten-out cross validation, we randomly split the data set into two non-overlapping sets (training vs. testing) forty times. Each time, the discriminative features are re-learned from the 30 MR images in the training set only. Then the learned top-N features (biomarkers) are used to classify the 10 unseen images. The mean and the variance of the sensitivity and specificity of the test images are computed and reported over the total number (40 times) of the random splits.

We use the K-nearest neighbor classifier (KNN) [12], with K varying from 1 to 25, in the top-N most discriminative feature space with N varying from 3 to 100 (Table 1). As a comparison, we also show the rates of using the Hippocampus volume of the same data (interactively segmented by human experts) as the discriminative measurements (bottom of Table 1).

Table 1. Leave-Ten-Out Cross Validation Results using KNN

Feature	Top-N Features	LTO Sensitivity mean(std)	LTO Specificity mean(std)
Features	3	88% (15%)	96% (0.7%)
Extracted from	5	88% (15%)	97% (0.5%)
Deformation & Tensor Fields	50	91% (12%)	99% (0.3%)
	100	90% (14%)	98% (0.4%)
Hippocampus Volume (right)	1	71.8% (15.4%)	78.6%(7.9%)
Hippocampus Volume (left)	1	67.4%(21%)	68%(11.6%)

In addition, for the purpose of further understanding the effective combination of different types of non-redundant features and their anatomical relations, we carry out a feature subset selection process to find the best triplets in the top-N AVR-ranked features, with N varying from 30 to 100. We use the Linear Discriminant Analysis (LDA) Classifier and forward sequential feature subset selection strategy [12]. Consistent with our earlier work [13], more sophisticated classifiers do not provide better results. In each of the automatically selected triplet spaces, we are able to examine the distribution of the training and testing data, as well as how MCI subjects

are distributed in the discriminative feature spaces chosen for separating AD and normal controls (Figure 2).

Furthermore, to investigate the statistical stability of our results, we have also randomly permuted the labels of the MR images five times. Each time, perfect separations are achieved on the training data; however, the cross validation results on the previously unseen data are mean sensitivity and specificity = 50%, indicating that the system performs no better than chance as expected.

Figure 4 demonstrates that, contrary to the common believe of the danger of the curse of dimensionality and random, absurd features, the top-100 most discriminative features cluster in three distinct locations. These features concentrate on the critical areas in memory and cognitive processes. The ventral striatum has connections with the basal forebrain, which is the most vulnerable area in Alzheimer's disease, since it is the area where the cholinergic neurons are located, and are usually affected in the presymptomatic phase of Alzheimer's disease (Mesulam).

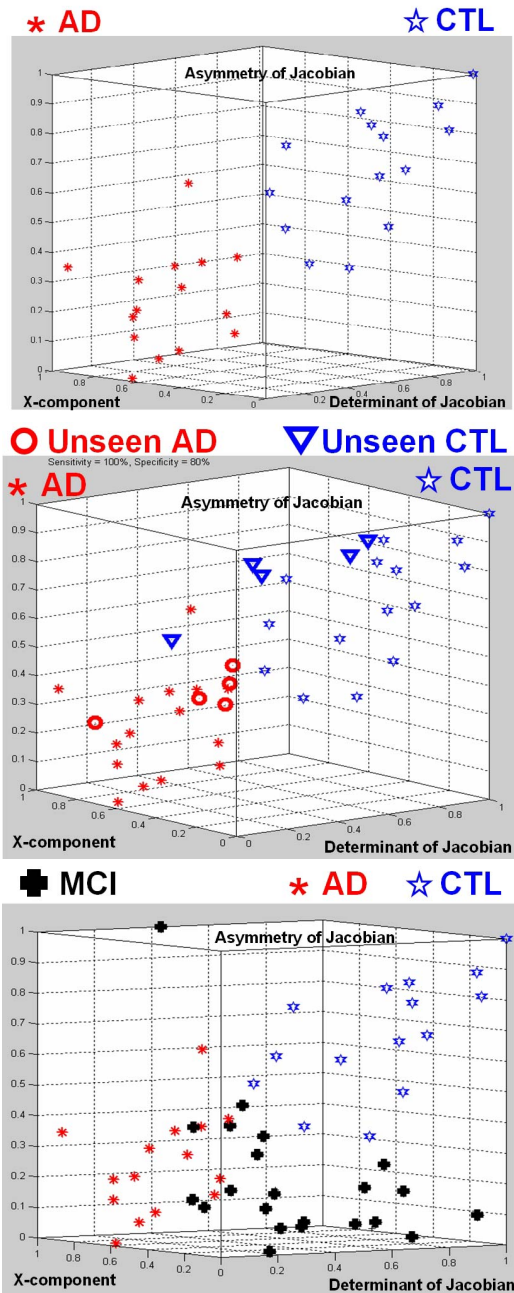
We run experiments on a 13-PC-cluster in parallel using non-optimized Matlab code. The cluster has 2.4 GHz Intel Xeon processors with 533 MHz front-side bus and 1GB of dual channel memory and an 80 Gb hard drive. It takes 7 hours for the registration of all data sets and 25 hours to generate all features (500 plus million). Thus there is a huge savings (e.g. 100/500 millions) in time during testing previously unseen MR images since we only need to compute those learned discriminative features on the test images. All time-consuming computation is done off-line, during the training stage only.

4. CONCLUSION

We have developed a computationally efficient approach for discovery of biomarkers of AD prediction from MR structural images. Our method has been tested stringently using leave-ten-out cross validations on unseen images with very promising results (> 90% sensitivity and specificity). Our results demonstrate the power of having a large image feature set combined with effective feature reduction tools, and simple classifiers to reduce a massive feature set to a low dimensional, practically useful, set of biomarkers for future MR image classification. The discriminative anatomical locations automatically found provide reassurance from literature and intriguing new evidence for further exploration and interpretation of biological causes of AD and paths of conversions from MCI to AD.

5. REFERENCES

- [1] R. Brookmeyer, S. Grey, and C. Kawas, "Projections of alzheimer's disease in the united states and the public health impact of delaying disease onset," *Am J Public Health*, vol. 88, pp. 1337-1342, 1998.
- [2] C. Davatzikos, M. Vaillant, S. Resnick, J.L. Prince, S. Letovsky, and R.N. Bryan, "A computerized approach for morphological analysis of the corpus callosum," *Comp. Ass. Tomography*, vol. 20, pp. 88-97, Jan./Feb. 1996.
- [3] A.M.C. Machado, J.C. Gee, and M.F.M. Campos, "Atlas warping for brain morphometry," *Proceedings of Medical Imaging, SPIE Press*, pp. 642-651, 1998.
- [4] C. Studholme, V. Cardenas, R. Blumenfeld, N. Schuff, H.J. Rosen, B. Miller, and M. Weiner, "Deformation tensor morphometry of semantic demnetia with quantitative validation," *NeuroImage*, vol. 22, pp. 1387-1398, 2004.
- [5] P.M. Thompson, W.P. Woods, M.S. Mega, and A.W. Toga, "Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain," *Human Brain Mapping*, vol. 9, pp. 81-92, 2000.
- [6] C. Davatzikos, "Whole-brain morphometric study of schizophrenia reveals a spatially complex set of focal abnormalities," *Archives of General Psychiatry*, 2005.
- [7] Y. Liu, R.T. Collins, and W.E. Rothfus, "Robust Mid-sagittal Plane Extraction from Normal and Pathological 3D Neuroradiology Images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 175-192, March 2001.
- [8] CJ Holmes, R Hoge, L Collins, R Woods, AW Toga, and AC Evans, "Enhancement of mr images using registration for signal averaging," *J Comput Assist Tomogr*, vol. 22, no. 2, pp. 324-33, Mar-Apr 1998.
- [9] J.C. Gee and R.K. Bajcsy, "Elastic matching: Continuum mechanical and probabilistic analysis," 1998.
- [10] J.P. Thirion, "Image matching as a diffusion process: An analogy with maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243-260, 1998.
- [11] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, "Itk software guide," *Kitware, Inc.*, 2003.
- [12] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.
- [13] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C.S. Carter, V.A. Stenger, S. Davis, H. Aizenstein, J. Becker, O. Lopez, and C. Meltzer, "Discriminative mr image feature analysis for automatic schizophrenia and alzheimer's disease classification," in *7th International Conference on Medical Imaging Computing and Comptuer Assisted Intervention (MICCAI 2004)*. October 2004, pp. 378-385, Springer.



(a) Result 1

Fig. 2. Top: the training set is separated completely by a set of discriminative triplet features for AD and normal control. Middle: when bringing in the ten previously unseen samples, nine out of the ten are correctly classified by a KNN classifier. Bottom: When project the 26 MCI subjects into this AD-CTL discriminative feature subspace they are spread out between and around AD and control subjects, indicating a potential converting process. It is interesting to notice that the normal subjects and the AD+MCI subjects have a significant difference along the brain asymmetry dimension (the vertical axis in the figure).

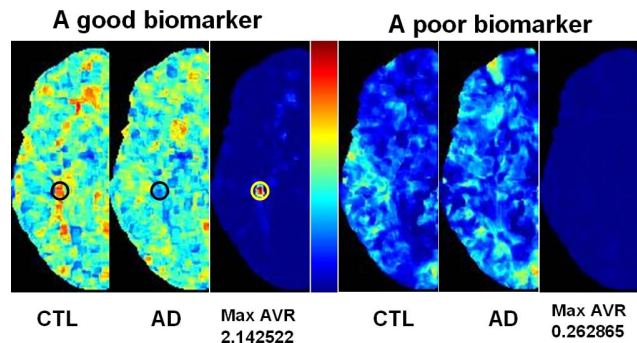


Fig. 3. Two different types of features are shown for AD and CTL respectively along with their voxel-wise AVR value maps (on slice 67 of the reference image): red — high AVR value, blue — low AVR value. The left feature type is the max value of the asymmetry of the tensor field (Jacobian of the deformation field) at scale 3 (finest scale) with neighborhood size $3 \times 3 \times 3$. The AVR values of this feature indicates that this is a good biomarker (location circled) for AD-CTL discrimination. The right feature type is the asymmetry measure of the θ angle value, moment 4, original resolution. Its AVR map indicates its poor discriminative power for AD-CTL.

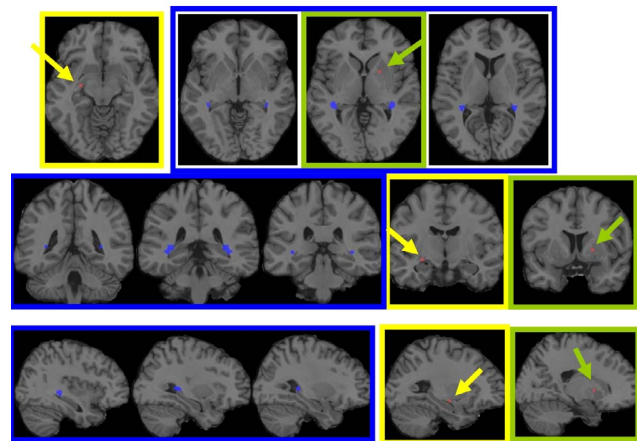


Fig. 4. The axial, coronal and sagittal views of the top-100 AVR-rated feature locations. These top-100 most discriminative features form three clusters: (1) the blue ones are the brain asymmetry measures in the Isthmus region, which is posterior to the stria terminalis. This part of the brain is the connection between the posterior temporal lobe and the cingulate gyrus; (2) the pink feature in the left side of the axial and coronal scans is the right Uncinate Fasciculus. It connects the anterior temporal lobe with the inferior frontal lobe; and (3) the pink feature in the right side of the axial and coronal scans is the Left ventral striatum. These are critical areas in memory and cognitive processes.