Quantization and analysis of hippocampal morphometric changes due to dementia of Alzheimer type using metric distances based on large deformation diffeomorphic metric mapping

Elvan Ceyhan a,∗, Mirza Faisal Beg b, Can Ceritoğlu c, Lei Wang d,e, John C. Morris f,g, John G. Csernansky d, Michael I. Miller c,h, J.Tilak Ratnanather c,h

ARTICLE INFO

Article history:
Received 27 October 2009
Received in revised form 2 September 2010
Accepted 21 January 2011

Keywords:
Computational anatomy
Hippocampus
Left–right asymmetry
Logistic discrimination
Morphometry
Repeated-measures ANOVA

ABSTRACT

The metric distance obtained from the large deformation diffeomorphic metric mapping (LDDMM) algorithm is used to quantize changes in morphometry of brain structures due to neuropsychiatric diseases. For illustrative purposes we consider changes in hippocampal morphometry (shape and size) due to very mild dementia of the Alzheimer type (DAT). LDDMM, which was previously used to calculate dense one-to-one correspondence vector fields between hippocampal shapes, measures the morphometric differences with respect to a template hippocampus by assigning metric distances on the space of anatomical images thereby allowing for direct comparison of morphometric differences. We characterize what information the metric distances provide in terms of size and shape given the hippocampal, brain and intracranial volumes. We demonstrate that metric distance is a measure of morphometry (i.e., shape and size) but mostly a measure of shape, while volume is mostly a measure of size. Moreover, we show how metric distances can be used in cross-sectional, longitudinal analysis, as well as left–right asymmetry comparisons, and provide how the metric distances can serve as a discriminative tool using logistic regression. Thus, we show that metric distances with respect to a template computed via LDDMM can be a powerful tool in detecting differences in shape.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Methods developed in the field of computational anatomy (CA) that enable quantification of brain structure volumes and shapes between and within groups of individuals with and without various neurological diseases have emerged from several groups in recent years [1–6]. Based on the mathematical principles of general pattern theory [7,8,9,10], these methods combine image-based diffeomorphic maps between MR scans with representations of brain structures as smooth manifolds. Because of their high repeatability and sensitivity to changes in neuroanatomical shapes, they can be especially sensitive to abnormalities of brain structures associated with a disease. Numerous post-mortem studies have shown that hippocampus is implicated with dementia of the Alzheimer type (DAT) [11–17]. The accumulation of neurofibrillary tangles and amyloid plaques characteristic of AD are associated with neuronal damage and death [18]. Furthermore, macroscopic gray matter losses from the accumulation of microscopic scale neuronal destruction are detectable in living subjects using currently available magnetic resonance (MR) imaging. Specifically, volume losses within the hippocampus [19–24] have recently been reported in subjects with mild-to-moderate AD. Using image-based diffeomorphic maps, we previously demonstrated that the combined assessment of hippocampal volume loss and shape deformity optimally distinguished subjects with very mild DAT from both elder nondemented subjects and younger healthy subjects [20]. These methods also allowed us to demonstrate that hippocampal shape deformities associated with very mild DAT and nondemented aging were distinct [25]. These methods were also extended to quantify changes in neuroanatomical volumes and shapes within...
the same individuals over time [26]. Other longitudinal neuroimaging analysis of hippocampal structures in individuals with AD have also emerged [27–38].

An important task in CA is the study of neuroanatomical variability. The anatomic model is a quadruple \((\Omega, \mathcal{G}, \mathcal{I}, \mathcal{T})\) consisting of \(\Omega\) the template coordinate space (in \(\mathbb{R}^3\)), defined as the union of 0, 1, 2, and 3-dimensional manifolds, \(\mathcal{G}: \Omega \rightarrow \Omega\) a set of diffeomorphic transformations on \(\Omega\), \(\mathcal{I}\) the space of anatomies is the orbit of a template anatomy \(I_0\) under \(\mathcal{G}\), and \(\mathcal{T}\) the family of probability measures on \(\mathcal{G}\). In this framework, a geodesic \(\phi: [0, 1] \rightarrow \mathcal{G}\) is computed where each point \(\phi_t = \phi(t) \in \mathcal{G}, t \in [0, 1]\) is a diffeomorphism in the domain \(\Omega\). The evolution of the template image \(I_0\) along path is given by \(I_t = \phi^{-1}_t I_t\) such that the end point of the geodesic connects the template \(I_0\) to the target \(I_1\) via \(I_t = \phi_t I_0 = I_0 \circ \phi_t^{-1}\). Thus, anatomical variability in the target is encoded by these geodesic transformations when a template is fixed. Furthermore, geodesic curves induce metric distances between the template and the target shapes in the orbit as follows. The diffeomorphisms are constructed as a flow of ordinary differential equations \(\dot{\phi}_t = \nu(t)\phi_t\), \(t \in [0, 1]\) with \(\phi_0 = \text{id}\) the identity map, and associated vector fields, \(v_t, t \in [0, 1]\). The optimal velocity vector field parameterizing the geodesic path is found by solving

\[
\nu = \arg\min_{\nu} \int_0^1 \|v_t\|_V^2 \, dt \quad \text{such that } I_0 \circ \phi_t^{-1} = I_1,
\]

where \(v_t \in V\), the Hilbert space of smooth vector fields with norm \(\|\cdot\|_V\) defined through a differential operator enforcing smoothness [26]. The length of the minimal length path through the space of transformations connecting the given anatomical configurations in \(I_0\) and \(I_1\) defines a metric distance between anatomical shapes in \(I_0\) and \(I_1\) via

\[
d(l_0, l_1) = \int_0^1 \|\tilde{v}_t\|_V \, dt,
\]

where \(\tilde{v}_t\) is the optimizer calculated from the large deformation diffeomorphic metric mapping (LDDMM) algorithm [39]. Here, the metric distance does not have any units. The construction of such a metric space allows one to quantify similarities and differences between anatomical shapes in the orbit. This is the vision laid out by D’Arcy W. Thompson almost one hundred years ago. Figure 1 exemplifies the change in the metric distance during the evolution of the diffeomorphic map from the template shape to the target shape.

The notion of mathematical biomarker in the form of metric distance can be used in different ways. One is to generate metric distances of shapes relative to a template [39,40]. Another is to generate metric distances between each shape within a collection [41]. The latter approach allows for sophisticated pattern classification analysis; it is however computationally expensive. We present an analysis based on the former approach which could provide a powerful tool in analyzing subtle shape changes over time with considerably less computational load. This approach may allow detecting the subtle morphometric changes observed in the hippocampus in DAT subjects in particular for those previously analyzed [26,42]. These studies compared rates of change in hippocampal volume and shape in subjects with very mild DAT and matched (for age and gender) nondemented subjects. The change in hippocampal shape over time was defined as a residual vector field resulting from rigid-body motion registration, and changes in patterns along hippocampal surfaces were visualized and analyzed via a statistical measure of individual and group change in hippocampal shape over time and used to distinguish between the subject groups. Hence the motivation to analyze LDDMM generated metric distances between binary hippocampus images at baseline and at follow-up with respect to the same template hippocampus image. That is, the template was compared again, and not propagated between time points. One might wonder why we do not track changes within a subject directly, rather than via a reference template, as it could give a more sensitive measure of shape change since the small difference in shape would make finding correspondence more accurate. Although we have considered doing this, the difficulty is that since the template (or origin) is different for each longitudinal computation, how to correctly perform statistical comparison of group change is not completely settled. This is actively being developed by using the concept of “parallel transport” [43,44].

In this study, we compute and characterize what metric distances convey and provide a statistical methodology to analyze metric distances based on the data used in [26]. In fact, [26] used a previous implementation of the present technology called large-deformation high-dimensional brain mapping (or HDBM-LD). The method was also based on use of a single template and large deformation diffeomorphic mappings. The differences between the methodology of [26] and this article is the way in which the resulting deformation fields that represent morphometric differences were analyzed. For the former was analyzed using singular value decomposition (SVD) applied on each incremental deformation field; the present is analyzed according to only its metric shape difference.

The morphometric differences are measured with respect to a template, so the differences in morphometry are relative to this template. However, such differences might also imply the correlation of morphometry with certain types of conditions. For example, a cross-sectional analysis of distances for different conditions or groups might suggest changes in morphometry due to a disease. Longitudinal analysis of distances might provide how the morphometry changes over time and whether the change is different between the groups. Under certain restrictions, even left–right anatomical asymmetry may be measured and analyzed for different groups. Moreover, metric distances (together with some other measures such as volume) might also be used for discriminative purposes; i.e., they can be used as biomarkers to distinguish healthy tissues from diseased ones.

We briefly describe computation of metric distances via LDDMM in Section 2.1, the hippocampal data set in Section 2.2, statistical methods we employ in Section 2.3. Furthermore, we provide the analysis on metric distances and volumes in Section 3, compare volumes and metric distances in Section 3.8, and analyze annual percentage rate of change in volumes and distances in Section 4. In the final section, we discuss the use of metric distances for baseline-followup studies, group comparisons, and discrimination analysis.

### 2. Methods

#### 2.1. Computing metric distance via large deformation diffeomorphic metric mapping

Metric distances between the binary images and the template image are obtained by computing diffeomorphisms between the images. Computation and analysis of these diffeomorphic mappings have been previously described [45]. Diffeomorphisms are estimated via the variational problem that, in the space of smooth velocity vector fields \(V\) on domain \(\Omega\), takes the form [39]:

\[
\nu = \arg\min_{\nu: \phi_t = v_t(\phi_t)} \left( \int_0^1 \|v_t\|_V^2 \, dt + \frac{1}{\sigma^2} \|I_0 \circ \phi_t^{-1} - I_1\|_G^2 \right).
\]
The optimizer of this cost generates the optimal change of coordinates $\psi = \phi^t \circ \phi^{-1}$ upon integration $d\phi_{\psi}^t / dt = d\psi / dt$, $\phi_0 = id$, where the subscript $\nu$ in $d\phi_{\psi}^t / dt$ is used to explicitly denote the dependence of $\phi$ on the associated velocity field $v$. Enforcing a sufficient amount of smoothness on the elements of the space $V$ of allowable velocity vector fields ensures that the solution to the differential equation $\dot{\phi}_t = v(t, \phi_t)$, $t \in [0, 1]$, $\nu \in V$ is in the space of diffeomorphisms $\mathcal{D}([0, 1], \mathcal{K})$. The required smoothness is enforced by defining the norm on the space $V$ of smooth velocity vector fields through a differential operator $L$ of the type $L(\alpha t + \gamma, \alpha t)_{\nu}$, where $\alpha > 1.5$ in $3$-dimensional space such that $||L\nu||_2 = ||L\nu||_1$, and $||\cdot||_2$ is the standard $L_2$ norm for square integrable functions defined on $\Omega$. The gradient of this cost is given by

$$
\nabla_v E_t = 2\nu - K \left( \frac{2}{\sigma^2} |D\phi_{\psi}^t| \nabla |f - I^t| \right)
$$

where $f^t = id \circ \phi_t$ and $I^t = I \circ \phi^{-1}$, $|Dg|$ is the determinant of the Jacobian matrix for $g$ and $K$ is a compact self-adjoint operator $K: L_2(\Omega, \mathbb{R}^d) \to V$ uniquely defined by $(a, b)_V = (Ka, b)_V$ such that for any smooth vector field $f \in V$, $Kf = f$ holds. The metric distance is then calculated via Eq. (2).

### 2.2. Subjects and data acquisition

Detailed description of subjects can be found in [26] where 18 very mild DAT subjects (Clinical Dementia Rating Scale, CDR0.5) and 26 age-matched nondemented controls (CDR0) were each scanned approximately two years apart. Clinical Dementia Rating (CDR) scale assessments which detect the severity of dementia symptoms were performed annually in all subjects by experienced clinicians without reference to neuropsychological tests or in vivo neuroimaging data. The experienced clinician conducted semi-structured interviews with an informant and the subject to assess the subject’s cognitive and functional performance: a neurological examination was also obtained. The clinician determined the presence or absence of dementia and, when present, its severity with the CDR. Overall CDR scores of 0 indicate no dementia, while CDR scores of 1, 2, and 3 indicate very mild, mild, moderate and severe dementia, respectively [48]. CDR assessments have been shown to have an inter-rater reliability of $k = .74$ (weighted kappa coefficient $k$ of 0.87) [50], and this high degree of inter-rater reliability has been confirmed in multi-center dementia studies [51]. Elderly subjects with no clinical evidence of dementia (i.e., CDR0 subject) have been confirmed with normal brains at autopsy with 80% accuracy; i.e., approximately 20% of such individuals show evidence of AD [52]. CDR0.5 subjects have subtle cognitive impairment, and 93% of them progress to more severe stages of illness (i.e., CDR > 1.5) and show neuropathological signs of AD at autopsy [53,54,52]. Although elsewhere the CDR0.5 individuals in our sample may be considered to have MCI [55], they fulfill our diagnostic criteria for very mild DAT and at autopsy overwhelmingly have neuropathologic AD [56]. A summary of subject information is listed in Table 1.

The scans were obtained using a Magnetom SP-4000 1.5 Tesla imaging system, a standard head coil, and a magnetization prepared rapid gradient echo (MPRAGE) sequence. The MPRAGE sequence (TR/TE = 10/4, ACQ = 1, Matrix = 256 x 256, Scanning time = 11.0 min) produced 3D data with a 1 mm x 1 mm in-plane resolution and 1 mm slice thickness across the entire cranium.

A neuroanatomical template was produced using an MR image from an additional elder control (i.e., CDR0 or non-demented) subject (male, age = 69). The choice and a detailed description of the template is provided in [45]. The subject selected to produce this template was obtained from the same source as the other subjects in the study, but was not otherwise included in the data analysis. Data used are the left and right hippocampal surfaces in the template scan created from expert–produced manual outlines using methods previously described [25,57], and the left and right hippocampal surfaces of each subject generated at baseline and follow-up. These surfaces were converted to binary hippocampus volumetric images by flood filling the inside of the surface and giving it label 1, and the outside of the surface was labeled as 0, or background. Each individual hippocampal surface was aligned with the template surface via a rigid-body rotation and translation before converting to volumetric binary images. In [57] we showed that mapping accuracy could be enhanced at higher resolution because of smaller voxels – voxels at the periphery of the structure (i.e., surface) account for much more of the structural volume at 1 mm$^3$ voxel resolution versus .5 mm$^3$. Since then we have adapted this as part of the standard mapping procedure. These surfaces were then converted into binarized image of dimension $64 \times 112 \times 64$ with voxel resolutions of $.5 \times .5 \times .5$ mm$^3$, followed by smoothing by a Gaussian filter of $9 \times 9 \times 9$-voxel window and one voxel standard deviation to smooth out the edges for LDDMM, which was then applied to each template-subject pair to compute metric distances, $d_k$, $d_k (k = 1, \ldots, 44)$, in each hemisphere at baseline (b) and at follow-up (f) as illustrated in Fig. 2. Segmentation of hippocampal MRI shapes across subjects, especially in diseased states, is a challenging problem. However the accuracy of the segmentation is not the point of this paper and has been demonstrated before [20,45,57].

In addition to the metric distances, our data set also consists of the following variables: gender, age, education in years (these variables are used for controlling the confounding affects of these factors on hippocampus morphometry). Furthermore, we have brain and intracranial volumes at baseline and followup, and hippocampus volumes for left and right hippocampi at baseline and follow-up. For the left data, hippocampi at baseline for CDR0.5 subjects are labeled as LB-CDR0.5, at follow-up they are labeled as LF-CDR0.5. CDR0 hippocampi are labeled as LB-CDR0 and LF-CDR0 accordingly. Similar labeling is done for the right metric distances.

### 2.3. Statistical methods

First, we investigate what metric distance measures and how it is related to hippocampal, brain, and intracranial volumes. That is, as a compound measure of morphometry, how much of the metric distance is related to shape and volume. Along this line, we provide the correlation between volume and metric distance measures by the pairs plots at baseline and follow-up of left and right hippocampi. Furthermore, we perform a principal component analysis

---

**Fig. 1.** Change in metric distance during diffeomorphic flow from template ($I_0$) to target ($I_1 = \phi_1 I_0 = b_0 \circ \phi_1^{-1}$). The numbers are the metric distance estimates at the three intermediate stages and the final stage of the LDDMM algorithm.
(PCA) on metric distance and volumes to characterize the major traits these quantities measure [58,59]. Since the metric distance between two images is just a single number, the PCA involves only the relationships among metric distance, volume, follow-up interval, and functional measurements.

We also provide a statistical methodology for the analysis of metric distances. We apply repeated-measures analysis of variance (ANOVA) on metric distances with diagnosis group as the main effect factor and side × timepoint as the repeated factor, since there is within-subject dependence of metric distances for left and right hemispheres and at baseline and follow-up. With this set-up we can both perform cross-sectional and longitudinal comparisons with metric distances. We apply four possible competing models each assuming a different variance-covariance structure to obtain the model that best fit our data set. The first model assumes compound symmetry, in which the diagonals (i.e., the variances) are equal, and so are the off diagonals (i.e., the covariances). The other three models assume unstructured, autoregressive (AR), and autoregressive heterogeneous variances, respectively. In the unstructured model, each variance and covariance term is different, in the AR model, the variances are assumed to be equal but the covariances change by time, and in the ARH model, the variances are also different and the covariances change by time [60–62]. We use various model selection criteria (such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood) to compare competing models to see which model best fits our data [63].

For post hoc comparison of the metric distances for baseline left, baseline right, and follow-up left, follow-up right hippocampi, we use Wilcoxon test and Brown and Forsythe’s (B–F) test (i.e., Levene’s test with absolute deviations from the median) for homogeneity of the variances [64]. The metric distances of the same person’s

![Fig. 2. Generation of metric distances $d_{k}^{b/f}$ for subjects $k = 1, \ldots, 44$ at baseline ($b$) and at follow-up ($f$).]
hippocampus at baseline and follow-up are dependent, hence we use Wilcoxon signed rank test to compare them. Likewise, the distances for the left and right hippocampi of the same person are also dependent. We estimate the empirical cumulative distribution functions (cdf) of the metric distances and compare them by Kolmogorov–Smirnov (K–S) test.

We apply logistic discrimination with metric distances and other variables, since the diagnosis have only two levels, namely CDR0 and CDR0.5. We use logistic regression to estimate or predict the risk or probability of having DAT using metric distances, together with side (i.e., hemisphere) and timepoint (baseline versus follow-up) factors. In other words, we model the probability that the subject is CDR0.5 given the metric distance of the subject for left or right hippocampus at baseline or follow-up. In standard logistic regression the model-parameters are obtained via maximum likelihood estimators. For more on logistic regression and logistic discrimination, see [65] and [58], respectively. First we model with one predictor variable at a time from side, timepoint, and metric distance, etc., if the variable is not significant at .05 level, we omit that variable from further consideration. We consider the full logistic model with the response logit \( p = \log \left( \frac{1}{1 - p} \right) \) where \( p = P(Y = 1) \) (i.e., the probability that the subject is diagnosed with CDR0.5) with one predictor variable at a time from side, timepoint, and metric distance, etc., if the variable is not significant at .05 level, we omit that variable from further consideration. The remaining variables with all possible interactions among them as the predictor variables constitute the full model. On this full model, we choose a reduced model by AIC in a stepwise algorithm and then use stepwise backward elimination procedure on the resulting model [63]. We stop the elimination procedure when all the remaining variables are significant at \( \alpha = .05 \) level. Based on the final model with significant predictors, we apply logistic discrimination. Notice that the subjects are evenly distributed in terms of gender, years of education, scan intervals, and age between the diagnostic groups so that these variables are accounted for as covariates. The brain and intracranial volumes are much larger in scale than the hippocampal volumes which are larger than the metric distances. Notice that brain and hippocampal volumes all decrease by time and are smaller in CDR0.5 subjects compared to CDR0 subjects. On the other hand, the metric distances tend to increase by time and are larger for the CDR0.5 subjects. Also presented in Table 1 are the \( p \) values for Lilliefors’ test of normality (\( p_L \)) and Wilcoxon rank sum test for differences between the diagnostic groups (\( p_W \)). Notice that most variables can be assumed to follow a Gaussian distribution, but since a few of the variables are non-Gaussian, we apply the Wilcoxon rank sum test instead of Welch’s \( t \)-test. The diagnostic groups do not significantly differ in age, education, brain and intracranial volumes. Furthermore, among the metric distances, we see that only right follow-up metric distances are significantly different between the diagnostic groups.

We present the pairs plot (scatter plot of each pair) of continuous variables in Fig. 3 and calculate the correlation coefficients between each pair of the variables (not presented). We observe that age and education are not significantly correlated with any of the other variables. Hence we discard them in our prospective analysis (except for logistic discrimination). We observe significant correlation between each pair of hippocampal volumes, and between each pair of brain and intracranial volumes. The metric distances are only moderately correlated with each other. Hippocampal volumes are mildly correlated with brain and intracranial volumes. The same holds for the metric distances but to a lesser extent.

See Fig. 4 for the (jittered) scatter plots of the metric distances by group, where the crosses are centered at the mean distances and the points are jittered (scattered) along the horizontal axis in order to avoid frequent point concurrence and tight clustering of points, thereby making the plot better for visualization. Observe that CDR0 distances are smaller than CDR0.5 distances at baseline and at follow-up for both left and right hippocampi. This suggests that the morphometric differences of CDR0 hippocampi with respect to the template hippocampus are smaller than those of CDR0.5.
Fig. 3. Pairs plots of the continuous variables for the hippocampi at baseline and follow-up. HLV: volume of left hippocampus; HRV: volume of right hippocampus; HLM: metric distance for left hippocampus; HRM: metric distance for right hippocampus; BV: brain volume; ICV: intracranial volume. The numbers 1 and 3 stand for year 1 (i.e., baseline) and year 3 (i.e., follow-up), respectively.

hippocampi. This is not surprising, considering the template hippocampus being one of the CDR0 hippocampi. Furthermore, the standard deviations of the distances for CDR0 subjects tend to be smaller than those of CDR0.5 subjects. That is, the morphometric variability of CDR0 hippocampi with respect to the template hippocampus is smaller than that of CDR0.5 hippocampi. The statistical significance of these results will be provided in the following sections.

3.2. Principal component analysis of the volumes and metric distances

The volumes and metric distances measure different but related aspects of morphometry, so some of the variables are highly correlated with each other (see Fig. 3). We perform principal component analysis (PCA) to obtain a set of uncorrelated variables that hopefully represent some identifiable aspect of the morphometry [58, 59]. In [26], a PCA in the form of singular value decomposition (SVD) was applied to the matrix of change in the vector fields (i.e., matrix of voxel displacements as a function of original coordinates), which is also obtained from the LDDMM algorithm. However, in this article, we perform the PCA procedure on hippocampal volume, metric distance, brain volume, and intracranial volume measures. When PCA (with eigenvalues based on the covariance matrix) on metric distances and volumes of left hippocampi at baseline is performed, we observe that the first principal component (PC1) accounts for almost all the variation (see Table 2). Considering the variable loadings, PC1 seems to be the head size component, PC2 is the contrast between brain and intracranial volumes, PC3 is the hippocampus size, and PC4 is the metric distance compo-
The variable loadings for the variables suggest that brain and intracranial volumes are mostly measures of head size, metric distance is mostly a measure of hippocampus shape and partly related to head and hippocampus sizes, and hippocampus volume is mostly a measure of hippocampus size and partly related to hippocampus shape and head size. Hence, one should use volume and metric distance in morphometric analysis of brain tissues as they provide complementary information.

3.3. Repeated-measures analysis of metric distances and hippocampal volumes

Due to within-subject dependence of metric distances for left and right hemispheres and for baseline and follow-up measures, we apply repeated-measures analysis with group or side as main factors. However, the volumes are in mm$^3$ and metric distances are unitless, hence the data are not to scale. In particular, the brain and intracranial volumes have the largest variation in the data set, hence dominate the PCs. To remove the influence of the scale (or unit), we apply PCA with eigenvalues based on the correlation matrix (i.e., PCA on the standardized variables). The importance scores of principal components and variable loadings from the PCA of metric distances and volumes of left hippocampi at baseline with eigenvalues based on the correlation matrix are presented in Table 3. Notice that with the correlation matrix, the first three PCs account for almost all the variation in the variables. Comparing the variable loadings, PC1 seems to be the head size component, PC2 is the hippocampus shape, PC3 is the hippocampus size and the contrast between hippocampus and head size, and PC4 is the contrast between brain and intracranial volume. The PCA on variables for left hippocampi at followup and right hippocampi at baseline and followup yields similar results (see [68]).

The variable loadings for the variables suggest that brain and intracranial volumes are mostly measures of head size, metric distance is mostly a measure of hippocampus shape and partly related to head and hippocampus sizes, and hippocampus volume is mostly a measure of hippocampus size and partly related to hippocampus shape and head size. Hence, one should use volume and metric distance in morphometric analysis of brain tissues as they provide complementary information.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
</table>

The importance of principal components and variable loadings from the principal component analysis of metric distances and volumes of left hippocampi at baseline with eigenvalues based on the covariance matrix. PCi stands for principal component i for i = 1, 2, 3, 4; Prop.Var: proportion of variance explained by the principal components; Cum.Prop: cumulative proportion of the variance explained by the particular principal component; HLV: volume of left hippocampus at baseline; HLM: metric distance of left hippocampus at baseline; BV: brain volume at baseline; ICV: intracranial volume at baseline.

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance of components</td>
<td>Prop. Var</td>
<td>.9877</td>
<td>.0123</td>
</tr>
<tr>
<td>Cum. Prop</td>
<td>.9877</td>
<td>~1.0</td>
<td>~1.0</td>
</tr>
<tr>
<td>Variable loadings</td>
<td>HLV</td>
<td>~.0</td>
<td>~.0</td>
</tr>
<tr>
<td>HLM</td>
<td>~.0</td>
<td>~.0</td>
<td>~.0</td>
</tr>
<tr>
<td>BV</td>
<td>.55</td>
<td>~.83</td>
<td>~.0</td>
</tr>
<tr>
<td>ICV</td>
<td>.83</td>
<td>.55</td>
<td>~.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
</table>

The importance of principal components and variable loadings from the principal component analysis of metric distances and volumes of left hippocampi at baseline and followup with eigenvalues based on the correlation matrix. HLV: volume of left hippocampus; HLM: metric distance of left hippocampus; BV: brain volume; ICV: intracranial volume. The other abbreviations are as in Table 2.

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Followup</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance of components</td>
<td>Prop. Var</td>
<td>.57</td>
<td>.27</td>
<td>.15</td>
<td>.01</td>
<td>.54</td>
<td>.32</td>
<td>.12</td>
<td>.02</td>
</tr>
<tr>
<td>Cum. Prop</td>
<td>.57</td>
<td>.84</td>
<td>.99</td>
<td>1.0</td>
<td>.54</td>
<td>.86</td>
<td>.98</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Variable loadings</td>
<td>HLV</td>
<td>.37</td>
<td>.59</td>
<td>.71</td>
<td>~.0</td>
<td>.41</td>
<td>.55</td>
<td>.73</td>
<td>~.0</td>
</tr>
<tr>
<td>HLM</td>
<td>~.22</td>
<td>~.80</td>
<td>.55</td>
<td>~.0</td>
<td>~.0</td>
<td>~.80</td>
<td>.60</td>
<td>~.0</td>
<td></td>
</tr>
<tr>
<td>BV</td>
<td>.64</td>
<td>~.0</td>
<td>~.27</td>
<td>~.72</td>
<td>.65</td>
<td>~.16</td>
<td>~.18</td>
<td>~.72</td>
<td></td>
</tr>
<tr>
<td>ICV</td>
<td>.63</td>
<td>~.0</td>
<td>~.33</td>
<td>.70</td>
<td>.64</td>
<td>~.17</td>
<td>~.29</td>
<td>.69</td>
<td></td>
</tr>
</tbody>
</table>
effects and timepoint as the repeated factor (see [68]), and group as main effect and side \times timepoint as the repeated factor (see below). We have four measurements for each subject, namely left metric distance at baseline, left metric distance at followup, right metric distance at baseline, and right metric distance at followup, so repeated-measures analysis can be performed on our data set.

We model the Var-Cov structure for the repeated-measures for each subject. We have four correlated measures per subject, namely LDB, LDF, RDB, and RDF. We try compound symmetry, unstructured, autoregressive (AR), and autoregressive heterogeneous (ARH) Var-Cov structures. The variances (in the diagonal) suggest homogeneity, but covariances seem to differ. See [68] for the comparison of models. We choose the model with AR Var-Cov structure based on AIC and BIC. The corresponding model is

\[ d_{ijkl} = \mu + \alpha_i^S + \alpha_j^D + \alpha_k^T + \alpha_{ijkl}^ST + \epsilon_{ijkl}, \]

where \( d_{ijkl} \) is the distance for subject i for side i (1 for left; 2 for right) with diagnosis j (j = 1 for CDR0; 2 for CDR0.5) at timepoint k (k = 1 for baseline; 2 for follow-up), \( \mu \) is the overall mean, \( \alpha_i^S \) is the effect of side level i, \( \alpha_j^D \) is the effect of diagnosis level j, \( \alpha_k^T \) is the effect of timepoint level k, \( \alpha_{ijkl}^ST \) is the side \times diagnosis interaction, \( \alpha_{ijkl}^T \) is the timepoint \times diagnosis interaction, \( \alpha_{ijkl}^ST \) is the side \times diagnosis \times timepoint interaction, and \( \epsilon_{ijkl} \) is the error term. The Var-Cov structure for the error term (for the four measures per subject) is

\[ \text{Cov}(\epsilon_{ijkl}, \epsilon_{ijkl'}) = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 & \sigma_1 \sigma_3 & \sigma_1 \sigma_4 \\ \sigma_1 \sigma_2 & \sigma_2^2 & \sigma_2 \sigma_3 & \sigma_2 \sigma_4 \\ \sigma_1 \sigma_3 & \sigma_2 \sigma_3 & \sigma_3^2 & \sigma_3 \sigma_4 \\ \sigma_1 \sigma_4 & \sigma_2 \sigma_4 & \sigma_3 \sigma_4 & \sigma_4^2 \end{bmatrix}. \]

The three way interaction of side \times group \times timepoint is not significant (\( F = 0.50, df = 1, 168, p = 0.4823 \)), and neither are the two way side \times group (\( F = 7.6, df = 1, 168, p = 0.3860 \)) and side \times timepoint interactions (\( F = 2.25, df = 1, 168, p = 0.1359 \)). On the other hand, the group \times timepoint interaction is significant (\( F = 4.87, df = 1, 168, p = 0.0041 \)). The main effects of side, group, and timepoint are all significant (\( F = 6.12, df = 1, 168, p = 0.0143; F = 4.05, df = 1, 168, p = 0.0457; \) and \( F = 19.52, df = 1, 168, p < 0.001 \), respectively), but due to interaction, the groups should be compared at each time point instead of an overall comparison of group means. So the resulting model with significant terms at \( \alpha = 0.05 \) level is

\[ d_{ijkl} = \mu + \alpha_i^S + \alpha_j^D + \alpha_k^T + \alpha_{ijkl}^ST + \epsilon_{ijkl}. \]

The interaction plots of diagnosis levels over the time-points for left and right hippocampi are presented in Fig. 5, where we observe that the lines are not parallel for the diagnosis which agrees with the significant group \times timepoint interaction. But, the main effects of timepoint and side being significant is interpretable between baseline and follow-up.

For modeling hippocampal volumes using the repeated-measures ANOVA with group as main effect and compound symmetry in Var-Cov structure and volume measurements repeated over time for each subject, and for modeling volumes using the repeated measures ANOVA with side as main effect and compound symmetry in Var-Cov structure and volume measurements repeated over time, see [68]. Volume measurements were also presented in detail in [26]. For the model that includes the diagnosis, side, and diagnosis \times side interaction, we find that the most promising model based on likelihood ratio test, BIC, and AIC is the one with unstructured Var-Cov matrix. The corresponding model with significant terms at \( \alpha = 0.05 \) level is

\[ V_{ijklm} = \mu + \alpha_i^S + \alpha_j^D + \alpha_k^T + \alpha_{ijkl}^ST + \alpha_{ijkl}^T + \alpha_{ijklm} + \epsilon_{ijklm}. \]
Fig. 5. Interaction plots for diagnosis levels over the timepoint levels for left and right metric distances demonstrating that the slopes are different between diagnostic groups (with slope differences in the right being larger).

In the post hoc comparison of hippocampal volumes, we find that the variances of volumes are not significantly different for (LB-CDR0.5, LB-CDR0), (RB-CDR0.5, RB-CDR0), and (RF-CDR0.5, RF-CDR0) groups, but volumes of LF-CDR0 hippocampi are significantly larger than volumes of LF-CDR0.5 hippocampi ($p = .0268$). The CDR0.5 volumes are significantly smaller than CDR0 volumes in left hippocampi at baseline ($p = .0001$) and follow-up ($p < .0001$), and for right hippocampi at baseline ($p = .0071$) and follow-up ($p = .0001$).

For the cdf comparisons, the samples (groups) should be independent for these tests to be valid, so we only compare CDR0.5 versus CDR0 groups. The results are provided in detail in [68] and for more on cdf comparisons, see [71,72]. The cdf of RF-CDR0.5 distances is significantly smaller than the cdf of RF-CDR0 distances ($p = .0259$ for K–S test). That is, RF-CDR0.5 metric distances are stochastically larger than RF-CDR0 right metric distances. In other words, RF-CDR0.5 hippocampus shapes are more likely to be different than the template hippocampus compared to RF-CDR0 hippocampus shapes. Furthermore, the cdf of LF-CDR0.5 distances is almost significantly smaller than the cdf of LF-CDR0 distances ($p = .0604$). Observe that these results are in agreement with the ones in Table 4. The CDR0.5 volumes are stochastically smaller than CDR0 volumes for left hippocampi at baseline ($p = .0007$) and follow-up ($p = .0003$), and for right hippocampi at baseline ($p = .0064$) and follow-up ($p = .0028$).

Remark 3.1. Although volume is a measure of size and metric distance is a measure of overall morphometric difference from a template, the repeated measure analysis and post hoc analysis of volumes and metric distances provide similar results. The main difference is that volumes tend to decrease, while metric distances tend to increase by time.

3.5. Comparison of baseline and follow-up metric distances and hippocampal volumes

For the comparison of dependent groups by paired difference method, see Table 4. Observe that LB-CDR0.5 metric distances are significantly smaller than LF-CDR0.5 distances at $\alpha = .05$ ($p = .0259$). The same holds for RB-CDR0.5 versus RF-CDR0.5 distances ($p = .0002$). That is, CDR0.5 hippocampi tend to become more different in morphometry from the template, which implies that for both left and right distances there is significant change in morphometry (perhaps reduction in size) of CDR0.5 hippocampi over time. In fact, significant volume reduction over time is detected

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The $p$-values based on Wilcoxon rank sum test for both left and right metric distances and volumes and $p$-values based on Wilcoxon signed rank test for metric distances and volumes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Groups</th>
<th>$p$-Values for $t$-test</th>
<th>$p$-Values for Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-sided</td>
<td>1st &lt; 2nd</td>
</tr>
<tr>
<td>LB-CDR0.5,LB-CDR0</td>
<td>.6078</td>
<td>.7044</td>
</tr>
<tr>
<td>LF-CDR0.5,LF-CDR0</td>
<td>.1625</td>
<td>.9223</td>
</tr>
<tr>
<td>RF-CDR0.5,RF-CDR0</td>
<td>&lt;.0212 *</td>
<td>.9900</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Groups</th>
<th>$p$-Values for paired $t$-test</th>
<th>$p$-Values for paired Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-sided</td>
<td>1st &lt; 2nd</td>
</tr>
<tr>
<td>LB-CDR0.5,LF-CDR0.5</td>
<td>.0311</td>
<td>.0155 *</td>
</tr>
<tr>
<td>RB-CDR0.5,RF-CDR0.5</td>
<td>.0005 *</td>
<td>.0002</td>
</tr>
<tr>
<td>RB-CDR0,RB-CDR0</td>
<td>.7127</td>
<td>.3563</td>
</tr>
<tr>
<td>LB-CDR0.5,RB-CDR0</td>
<td>.1244</td>
<td>.0622</td>
</tr>
<tr>
<td>LB-CDR0.5,RF-CDR0.5</td>
<td>.3465</td>
<td>.1733</td>
</tr>
<tr>
<td>LB-CDR0,RF-CDR0</td>
<td>.0385 *</td>
<td>.0192</td>
</tr>
<tr>
<td>LB-CDR0,RF-CDR0</td>
<td>.0585</td>
<td>.0292 *</td>
</tr>
</tbody>
</table>

* Significant $p$-values at .05 level.
The morphometric changes in CDR0.5 right hippocampi from baseline to follow-up are barely significantly larger than those of CDR0.5 left hippocampi \( (p = .0445) \). The associated \( p \)-value here is obtained by testing the difference sets \((LB-CDR0.5)-(LF-CDR0.5)\) versus \( (RB-CDR0)-(RF-CDR0.5) \) using the usual paired \( t \)-test. On the other hand, only RB-CDR0 is almost significantly less than RF-CDR0 \( (p = .0621) \), which implies there is some weak evidence for mild change in right hippocampi as a result of aging. Furthermore, the morphometric changes in CDR0 right hippocampi from baseline to follow-up are not significantly different from those of CDR0 left hippocampi \( (p = .3817) \).

The morphometric changes in CDR0.5 left hippocampi from baseline to follow-up are not significantly different from those of CDR0 left hippocampi \( (p = .1337) \), while the morphometric changes in CDR0.5 right hippocampi from baseline to follow-up are significantly larger from those of CDR0 right hippocampi \( (p = .0074) \). Therefore, over time, DAT influences the morphometry of right hippocampi more compared to left hippocampi.

The volumes decrease significantly by time in CDR0 group for both left and right hippocampi \( (p < .0001 \text{ for both}) \); the same holds for CDR0.5 group also \( (p = .0001 \text{ for both}) \). The volumetric reduction is significantly larger in CDR0.5 right hippocampi compared to CDR0.5 left hippocampi \( (p = .0407) \); but the same holds only barely in CDR0 group \( (p = .0524) \). On the other hand, the volumetric reduction is significantly larger in CDR0.5 left hippocampi compared to CDR0 left hippocampi \( (p = .0108) \); the same holds for right hippocampi also \( (p = .0418) \).

Such a baseline versus follow-up comparison is not performed in [26] neither in volumes nor in metric distances. The morphometric change in the subjects over time is analyzed based on the eigenfunctions obtained from the SVD procedure.

### 3.6. Comparison of left and right hippocampi

We also compare the presence of left–right morphometric asymmetry in hippocampi of healthy and diseased subjects. Such an asymmetry comparison is not performed in [26]. The left versus right comparisons are also presented in Table 4, where we see that LB-CDR0.5 and RB-CDR0.5 distances are not significantly different from each other \( (p = .3046) \), LF-CDR0.5 distances are significantly smaller than RF-CDR0.5 distances at .05 level \( (p = .0179) \), the same holds for LB-CDR0 versus RB-CDR0.5 \( (p = .0215) \) and LF-CDR0 vs RF-CDR0 \( (p = .0021) \) comparisons. This implies that at baseline morphometric differences of CDR0.5 left hippocampi from the left template are about the same as those of CDR0.5 right hippocampi from the right template. On the other hand at follow-up, morphometric differences of CDR0.5 left hippocampi are smaller than those of CDR0.5 right hippocampi at baseline and follow-up; morphometric differences of CDR0 left hippocampi from the left template are smaller than those of CDR0 right hippocampi, that is, CDR0 left hippocampi are more similar in morphometry to the left template when compared to CDR0 right hippocampi to the right template. These distance comparisons for left versus right hippocampi would imply left–right morphometric asymmetry, only if the left and right hippocampi of the template subject were same (up to a reflection). Otherwise, these comparisons are only suggestive of morphometric differences from the respective hemisphere (side) of the hippocampi.

The left hippocampus volumes are significantly smaller than the right hippocampus volumes at both baseline and follow-up years (i.e., there is significant volumetric left–right asymmetry in hippocampi); baseline volumes are larger than follow-up volumes for both left and right hippocampi (i.e., there is significant reduction in volume by time) \( (p < .0001 \text{ for each comparison}) \). We observe the same trend in the overall comparison for each group also. However, left–right volumetric asymmetry significantly reduces by time in CDR0.5 group \( (p = .0407) \); but the same holds only barely in CDR0 group \( (p = .0524) \). The level of left–right volumetric asymmetry is about the same in both CDR0 and CDR0.5 groups at baseline \( (p = .3495) \) and follow-up \( (p = .4853) \).

**Remark 3.2.** Correlation between Metric Distances of Dependent Hippocampi: Correlation coefficients between metric distances for baseline and follow-up (overall and by group) and for the left and right hippocampi are provided in [68]. Except for the CDR0 right hippocampi, the baseline and follow-up metric distances are significantly correlated for each group. That is, except for CDR0 right hippocampi, the distances tend to increase at baseline together with distances at follow-up. That is, as the morphometric differences from the template hippocampus increase at baseline, so do the differences from the template at follow-up (except for CDR0 right hippocampi). From the correlation analysis of left and right distances, we observe there is mild correlation between left and right metric distances. That is, as the morphometric differences of left hippocampi from the left template increase, differences of right hippocampi from the right template tend to increase slightly.

### 3.7. Logistic discrimination with metric distances and volumes

First we consider the full logistic model (designated as \( M_i(D) \)) with side, timepoint, and distance with all possible interactions being the predictor variables. In this model, each hippocampus MRI is treated as a distinct subject, hence we have \( 44 \times 4 = 176 \) values for each variable. When the stepwise model selection procedure is applied, the resulting model is

\[
\logit p_k = \beta_0 + \beta_1 d_{ik} + \beta_2 d_{ik}^2
\]

where \( \beta_2 \) is the coefficient of the quadratic term. In the logistic discrimination procedure, we can treat each hippocampus from left, right, baseline or follow-up hippocampi as a distinct subject. However, from a clinical point of view, each subject has four hippocampus MRIs in this study, and one MRI classified as CDR0.5 would suffice to classify the subject as CDR0.5, while all four MRIs should be classified as CDR0 for the subject to be classified as CDR0. Due to significant group \( \times \) timepoint interaction, we need to consider diagnosis groups at each time point. When we use baseline and follow-up distances one at a time in a logistic model, we will have \( 44 \times 2 = 88 \) values for each variable and we see that only the model

\[
\logit p_k = \beta_0 + \beta_1 d_{ik}^2 + \beta_2 (d_{ik}^2)^2
\]

has significant coefficients for the distance term. Again, labeling one of left or right hippocampi as CDR0.5 is sufficient to label the subject as CDR0.5 while both left and right MRIs should be classified as CDR0 for the subject to be classified as CDR0. Moreover, when we use baseline left, baseline right, follow-up left, and follow-up right distances (i.e., \( d_{i0L}, d_{i0R}, d_{i2L}, \) and \( d_{i2R} \)) one at a time in a logistic model, we will only have 44 values for each variable and we see that only the model

\[
\logit p_k = \beta_0 + \beta_1 d_{2R}^2
\]

has a significant coefficient for the distance term.
We apply the similar logistic discrimination methods on hippocampal volumes also. First we consider the full logistic model (designated as $M_0(V)$) with side, timepoint, and volume with all possible interactions being the predictor variables. After a stepwise elimination procedure we get the following reduced model:

$$M_{II}(V) : \logit p_i = \beta_0 + \alpha_k^{pl} + \beta_1 V_{ijkl}, \quad (11)$$

where $p_i$ is the probability of subject $i$ having DAT and $V_{ijkl}$ the volume for subject $i$ with diagnosis $j$ ($j=1$ for CDR0 and $2$ for CDR0.5) at timepoint $k$ ($k=1$ for baseline and $2$ for follow-up) with side $l$ ($l=1$ for left and $2$ for right), $\beta_0$ is the overall intercept, $\alpha_k^{pl}$ is the effect of side level $k$, and $\beta_1$ is the slope of the fitted line. When we use baseline and follow-up volumes (i.e., $V_{ijkl}$ and $V_{ijkl}^{pl}$) one at a time in a logistic model, we see that the model:

$$M_{II}(V) : \logit p_i = \beta_0 + \alpha_k^{pl} + \beta_1 V_{ijkl} \quad (12)$$

has the most significant coefficients for the volume terms. Moreover, when we use baseline left, baseline right, follow-up left, and follow-up right volumes (i.e., $V_{ijkl}^{pl}$, $V_{ijkl}^L$, $V_{ijkl}^R$, and $V_{ijkl}^{RF}$) one at a time in a logistic model, we see that the following model has the best fit:

$$M_{III}(V) : \logit p_i = \beta_0 + \beta_1 V_{ijkl}^{pl} \quad (13)$$

We could change the threshold probability in Eq. (5). The correct classification rates, sensitivity, and specificity percentages with $\pi_o \in \{1/2, 18/44\}$ are presented in Table 5. Observe that with $\pi_o = 1/2$ the best classifier among models $M_0(D) - M_0(D)$ is based on $M_0(D)$ and with $\pi_o = 18/44$ the best classifier is based on $M_0(D)$. Setting $\pi_o = 18/44$ (the proportion of CDR0.5 subjects in the data set) we get higher sensitivity rates than those with $\pi_o = 1/2$. Among models $M_0(V) - M_0(V)$, with $\pi_o = 1/2$ the best classifier is based on model $M_{II}(V)$ and with $\pi_o = 18/44$ the best classifier is based on model $M_{III}(V)$. However, as $\pi_o$ decreases, the correct classification rate and specificity tend to decrease.

One can optimize the threshold value of $\pi_o$ in Eq. (5) to maximize the correct classification rates and minimize the misclassification rates using an appropriately chosen cost function. For example one can consider the cost function

$$C_1(\pi_o, w_1, w_2) = -(\pi_{CDR0} - \pi_{CDR0.5}) w_1 (\pi_{CDR0.5} - \pi_{CDR0}) w_2, \quad (14)$$

where $w_1 < w_2$ are positive odd numbers, $F_{CDR0}$ is the number of CDR0.5 subjects classified (false) as CDR0 and $F_{CDR0.5}$ is the number of CDR0 subjects classified (false) as CDR0.5. Minimizing this cost function will maximize the correct classification rates and minimize the misclassification rates. Alternatively we can maximize the sensitivity and specificity rates by minimizing the following cost function

$$C_2(\pi_o, \eta_1, \eta_2) = - \left( \eta_1 \pi_{CDR0} - \pi_{CDR0.5} + \eta_2 \pi_{CDR0.5} - \pi_{CDR0} \right), \quad (15)$$

where $\eta_1, \eta_2 \geq 0$ and $\eta_1 + \eta_2 = 1$. Notice that as either of sensitivity or specificity increases, the cost function $C_2(\pi_o, \eta_1, \eta_2)$ decreases. If equal weight is put on the sensitivity and specificity, we can choose $w_1 = w_2 = 1$ in $C_1(\pi_o, w_1, w_2)$ and $\eta_1 = \eta_2 = 0.5$ in $C_2(\pi_o, \eta_1, \eta_2)$.

Using $w_1 = w_2 = 1$, optimal threshold values are $\pi_o = 0.5$ for model $M_II(D)$ in Eq. (9), $\pi_o = 0.45$ for model $M_III(D)$ in Eq. (10), and optimal $\pi_o = 0.38$ for model $M_III(V)$ in Eq. (11). The specificity rates are 69%, 73%, and 69%, respectively. The sensitivity rates are 56%, 67%, and 72%, respectively. Obviously, from a clinical point of view,
misclassifying a CDR0.5 subject as CDR0 (i.e., classifying a diseased subject as healthy) might be less desirable, since a subject labeled as CDR0.5 will undergo further screening but a subject labeled as CDR0 will be released. So the parameters $w_1$ and $w_2$ could be modified to reflect such practical concerns and then a different set of threshold $\pi_0$ values could be found. For example, we set $w_1 = 1$ and $w_2 = 3$ which favors correct classification of CDR0.5 subjects more than that of CDR0 subjects (i.e., favors higher specificity). Observe that with $w_1 = 1$ and $w_2 = 3$ the best classifier is based on model $M_{IV}(D)$ and with $w_1 = 3$ and $w_2 = 1$ the best classifier is based on model $M_{III}(D)$ with $\pi_0 = .3$. With $\pi_0 = .3$, $\hat{\beta}_1 = .7$ the best classifier is based on model $M_{III}(D)$, and with $\pi_0 = .3$, $\hat{\beta}_2 = .7$ the best classifier is based on model $M_{II}(D)$. Observe that from $\pi_0 = .3$ to $\pi_0 = .7$, sensitivity increases, correct classification rate and specificity tend to decrease.

The classification rates based on the logistic discrimination with hippocampal volumes are presented in Table 6. Observe that with $\pi_0 = 1/2$ the best classifier is based on model $M_{III}(V)$. Similarly, as $\pi_0$ decreases from $1/2$, sensitivity increases but the correct classification rate and specificity decrease. We use the cost function $C_2(\pi_0, w_1, w_2)$ with $w_1 = w_2 = 1$ and with $w_1 = 3$ and $w_2 = 3$ to calculate the optimal $\pi_0$ values for each of the models $M_{III}(V)$. Observe that with $w_1 = 3$ and $w_2 = 1$ the best classifier is based on model $M_{III}(V)$ and with $w_1 = 1$ and $w_2 = 3$ the best classifier is based on model $M_{II}(V)$. We find the optimal $\pi_0$ values based on the cost function $C_2(\pi_0, \eta_1, \eta_2)$ with $\pi_0 = .7$ and with $\eta_1 = .3$, $\eta_2 = .7$ for each of models $M_{III}(V)$. With $\pi_0 = .7$ the best classifier is based on model $M_{II}(V)$ and with $\eta_1 = .3$, $\eta_2 = .7$ the best classifier is based on model $M_{I}(V)$. Observe that from $\eta_1 = .7$ to $\eta_1 = .3$, $\eta_2 = .7$, sensitivity increases, correct classification rate and specificity tend to decrease.

3.8. Comparison of logistic discrimination procedures with hippocampal volumes and metric distances

Although volume is a measure of size and metric distance is a measure of overall morphometric difference from a template, the repeated measure analysis and post hoc analysis of volumes and metric distances provide similar results. The main difference is that volumes tend to decrease, while metric distances tend to increase by time. The logistic discrimination models are similar, except model $M_{IV}(D)$ for metric distances contains right follow-up distances, while model $M_{IV}(V)$ for volumes contains left follow-up volumes. The classification performances with $\pi_0 = 1/2$ and $\pi_0 = 18/44$ suggest that volume models have better performance than the metric distance models (see Tables 5 and 6). Using the optimal $\pi_0$ values with the cost functions $C_2(\pi_0, w_1, w_2)$ and $C_2(\pi_0, \eta_1, \eta_2)$, the classification performances are significantly different for models $M_3(V) - M_4(V)$ of volumes and $M_3(D) - M_4(D)$ metric distances. Comparing Tables 5 and 6, we see that logistic discrimination with volumes has better performance.

We apply the logistic discrimination using both volume and metric distance as predictors. The models we consider are the full logistic model (designated as model $M_3(V, D)$) with side, timepoint, volume, and metric distances with all possible interactions being predictor variables. We apply the same stepwise elimination procedure as in Section 3.7 and get

$$M_3(V, D) : \logit p_l = \beta_0 + \alpha^a_k + \beta_1 V_{ijkl} + \beta_2 d_{ijkl} + \beta_3 V_{ijkl} d_{ijkl}$$

where $p_l$ is the probability of subject $l$ having DAT and $V_{ijkl}$ the volume and $d_{ijkl}$ the distance for subject $l$ with diagnosis $i$ ($i = 1$ for CDR0 and $2$ for CDR0.5) at timepoint $j$ ($j = 1$ for baseline and $2$ for follow-up) with side $k$ ($k = 1$ for left and $2$ for right), $\beta_0$ is the overall intercept, $\alpha^a_k$ is the effect of side level $k$, $\beta_1$ is the coefficient for volume, $\beta_2$ is the coefficient for ninth power of the distance, $\beta_3$ is the coefficient for the interaction between volume and distance. When we use baseline or follow-up measures one at a time in a logistic model, we see that the model $M_3(V, D) : \logit p_l = \beta_0 + \alpha^a_k + \beta_1 V_{ijkl} + \beta_2 (d_{ijkl})^{9}$ has the most significant coefficients. When we use side $\times$ timepoint combinations one at a time in a logistic model, we see that the
following model has the best fit:

\[
M_V(V, D) : \logit p_i = \beta_0 + \beta_1 (d_{ij}^D)^3 + \beta_2 V_{ij}^L.
\]

The corresponding classification rates are presented in Table 7. Observe that considering metric distance and volume together in the logistic discrimination procedure with the cost functions \(C_1(\pi_a, w_1, w_2)\) and \(C_2(\pi_a, \eta_1, \eta_2)\), we get better classification rates compared to logistic models with only one of metric distance or volume being the predictors.

In [26], logistic discrimination procedure is applied with either on the first 12 eigenfunctions obtained from the SVD of vector field changes, or on these eigenfunctions together with the hippocampal volume changes. In this article, we perform logistic discrimination on volumes (hippocampal, brain and intracranial) and metric distances together with a model selection procedure. In the above logistic models, we consider the volumes only not the volume changes.

### 4. Annual percentage rates of change in hippocampal volumes and in metric distances

Our volume and metric distance comparisons are cross-sectional or longitudinal by construction. However these measures might need to be adjusted for anatomic variability, since intersubject variability might add substantial amount of noise to volume or distance measurements at baseline or follow-up. There is no simple way to correct for this noise in practice. Differential volume loss or distance change over time might be self-correcting for such variability. For example, entorhinal cortex volume loss over time was shown to be a better indicator for DAT than cross-sectional measurements [66].

The hippocampal volume change over time can be written as the following annual percentage rate of change (APC) [66]:

\[
V_{APC} = \frac{V_b - V_f}{V_b} \times T \times 100\%.
\]

where \(V_{APC}\) is the APC in volume for side \(j\) of subject \(k\) with diagnosis \(i\), \(\mu\) is the overall mean, \(d_{ij}^D\) is the effect of diagnosis level \(i\) (i.e., for CDRO; 2 for CDRO5), \(\psi^2_j\) is the effect of side level \(j\) (i.e., left and right), \(\epsilon_{ijk}\) is the error term. The diagnosis main effect is significant \((F = 18.62, df = 1, 84, p < .0001)\) but neither side main effect \((F = .72, df = 1, 84, p = .4754)\) nor diagnosis \(\times\) side interaction is significant \((F = .11, df = 1, 84, p = .7384)\). Consequently, we conclude that the lines that join the mean \(V_{APC}\) values in the interaction plot are parallel and far apart, the main effect of diagnosis comparison is meaningful, and about the same at each hemisphere. The post hoc comparison of \(V_{APC}\) indicates that the APC in CDRO5 volumes are significantly larger than APC in CDRO volumes \((p = .0001)\). The hippocampal metric distance change over time can be written as the following annual percentage rate of change:

\[
D_{APC} = \frac{d_{ik}^D - d_{ik}^F}{d_{ik}^D} \times T \times 100\%.
\]

where \(d_{ik}^D\) and \(d_{ik}^F\) are the distance changes. (18)

Notice that to make APC in metric distance positive, we take the difference \(d_{ik}^F - d_{ik}^D\) as opposed to the order in the APC in volume definition. However, since volume has a unit but metric distance is unitless, the same correction for anatomic variability might not work well for both measures. For metric distances, we recommend the measurement of the distance between follow-up and baseline hippocampi instead. In fact, the repeated-measures analysis of APC in metric distances do not perform better than the distances or APC in volumes [68].

### 4.1. Logistic discrimination based on annual percentage rates

We also apply the logistic discrimination methods of Section 3.7 on APC in hippocampal volumes. First we consider the full logistic model (designated as \(M_l(V_{APC})\)) with side and APC in volume with all possible interactions being the predictor variables. We apply the same stepwise elimination procedure as in Section 3.7 and get

\[
M_l(V_{APC}) : \logit p_k = \beta_0 + \beta_1 V_{APC,k} + \beta_2 (V_{APC,k})^3
\]

where \(p_k\) is the probability of subject \(k\) having DAT.

Furthermore, when we use \(V_{APC,L}\) and \(V_{APC,R}\) as predictors in a logistic model, we see that the following model has the best fit.

\[
M_{ll}(V_{APC}) : \logit p_k = \beta_0 + \beta_1 V_{APC,L,k} + \beta_2 V_{APC,R,k} + \beta_3 (V_{APC,L,k})^2
\]
The correct classification rates \( (P_{CR}) \), sensitivity \( (P_{Sens}) \), and specificity \( (P_{Spec}) \) percentages for the classification procedures based on models \( M(V, D) - M_{dc}(V, D) \) using hippocampal LDDMM metrics and volumes with optimum threshold values \( \pi_o = \pi_{opt} \) based on the cost function \( C_1(\pi_o, w_1, w_2) \) with \( w_1 = w_2 = 1 \) and \( w_1 = 1, w_2 = 3 \) (top); and with optimum threshold values \( \pi_o = \pi_{opt} \) based on the cost function \( C_2(\pi_o, \eta_1, \eta_2) \) with \( \eta_1 = \eta_2 = 5 \) and \( \eta_1 = 3, \eta_2 = 7 \) (bottom).

\[
\begin{array}{cccccc}
\pi_{opt} & 0.54–0.55 & 0.66 & 0.48–0.58 & 0.48 & 0.54–0.55 & 0.66 & 0.48–0.58 & 0.48 \\
P_{CR} & 84% & 84% & 84% & 84% & 84% & 84% & 84% & 84% \\
P_{Sens} & 89% & 83% & 83% & 83% & 89% & 83% & 83% & 89% \\
P_{Spec} & 81% & 85% & 69% & 85% & 81% & 85% & 69% & 73% \\
\end{array}
\]

Using optimum \( \pi_o \) based on cost function \( C_2(\pi_o, \eta_1, \eta_2) \) with

\[
\begin{array}{cccccc}
\pi_{opt} & 0.54–0.55 & 0.66 & 0.48–0.72 & 0.48 & 0.54–0.55 & 0.66 & 0.48–0.72 & 0.48 \\
P_{CR} & 84% & 84% & 80% & 84% & 84% & 84% & 80% & 84% \\
P_{Sens} & 89% & 83% & 61% & 83% & 89% & 83% & 100% & 94% \\
P_{Spec} & 81% & 85% & 92% & 85% & 81% & 85% & 50% & 61% \\
\end{array}
\]

The classification rates with \( \pi_o = 1/2 \) and \( \pi_o = 18/44 \) and optimal \( \pi_o \) values with respect to the cost functions are presented in Table 8. Observe that the classifier using the cost function \( C_2(\pi_o, \eta_1, \eta_2) \) with \( \eta_1 = 3, \eta_2 = 7 \) in model \( M(V^{APC}) \) has the best performance. Comparing Tables 7 and 8, we observe that correct classification rates, sensitivity, and specificity percentages with the classifiers based on APC in volume are about the same as those with volume only. Unlike the findings of [66] hippocampal volume loss over time is not a better indicator for DAT than cross-sectional measurements. On the other hand, the classifier based on volume and distance together performs better compared to models based on only one of volume, distance, or APC in volume values.

We apply the logistic discrimination methods of Section 3.7 on APC in hippocampal metric distances also in [68] where we demonstrate that hippocampal metric distance change over time is a poor indicator for DAT. We apply the logistic discrimination based on both the distance and APC in volume. First we consider the full logistic model (designated as \( M(V^{APC}, D) \)) with side and APC in volume, and distances with all possible interactions being the predictor variables. We apply the same stepwise elimination procedure as in Section 3.7 and get

\[
\log \text{it}_{p_k} = \beta_0 + \beta_1 V_{APC} + \beta_2 (V_{APC})^3 + \beta_3 d_{RF}^2. \quad (22)
\]

Furthermore, when we use left and right measures as predictors in a logistic model, we see that the following model has the best fit.

\[
\log \text{it}_{p_k} = \beta_0 + \beta_1 V_{APC} + \beta_2 (V_{APC})^3 + \beta_3 d_{RF}. \quad (21)
\]

The correct classification rates \( (P_{CR}) \), sensitivity \( (P_{Sens}) \), and specificity \( (P_{Spec}) \) percentages for the classification procedures based on models \( M(V^{APC}) - M_{dc}(V^{APC}) \) using APC in hippocampal volumes with optimum threshold values \( \pi_o = \pi_{opt} \) based on the cost function \( C_2(\pi_o, w_1, w_2) \) with \( w_1 = w_2 = 1 \) and \( w_1 = 1, w_2 = 3 \) (top); and with optimum threshold values \( \pi_o = \pi_{opt} \) based on the cost function \( C_2(\pi_o, \eta_1, \eta_2) \) with \( \eta_1 = \eta_2 = 5 \) and \( \eta_1 = 3, \eta_2 = 7 \) (bottom).

\[
\begin{array}{cccccc}
\pi_{opt} & 0.55–0.55 & 0.38–0.39 & 0.27–0.28 & 0.54–0.55 & 0.34 & 0.25 \\
P_{CR} & 80% & 75% & 75% & 80% & 73% & 73% \\
P_{Sens} & 72% & 72% & 78% & 72% & 83% & 83% \\
P_{Spec} & 85% & 81% & 73% & 85% & 65% & 65% \\
\end{array}
\]

Using optimum \( \pi_o \) based on cost function \( C_2(\pi_o, \eta_1, \eta_2) \) with

\[
\begin{array}{cccccc}
\pi_{opt} & 0.55–0.55 & 0.82–0.85 & 0.61–0.69 & 0.39 & 0.34 & 0.18–0.21 \\
P_{CR} & 80% & 82% & 82% & 64% & 73% & 70% \\
P_{Sens} & 72% & 56% & 56% & 89% & 83% & 100% \\
P_{Spec} & 85% & 100% & 100% & 46% & 65% & 50% \\
\end{array}
\]

* The model with the best classification performance.
observe that adding the metric distance to the logistic model with APC in volume improves the classification performance. Hence the model with hippocampal volume loss over time and metric distance is a better indicator for DAT compared to either variable used separately in logistic discrimination.

We also apply the logistic discrimination based on volume, distance, and APC in volumes. First, we consider the full logistic model (designated as $M(V, V^{APC}, D)$) with side, volume, and APC in volume, and distances with all possible interactions being the predictor variables. We apply the same stepwise elimination procedure as in Section 3.7 and get

$$M_{II}(V, V^{APC}, D) : \logit p_k = \beta_0 + \beta_1 V_{ik}^B + \beta_2 V_{ik}^{APC} + \beta_3 (V_{ik}^{APC})^3 + \beta_4 (d_{ik}^{B})^3.$$  

Furthermore, when we use left and right measures as predictors in a logistic model, we see that the following model has the best fit.

$$M_{III}(V, V^{APC}, D) : \logit p_k = \beta_0 + \beta_1 V_{ik}^{RF} + \beta_2 V_{ik}^{APC,R} + \beta_3 (V_{ik}^{APC,L})^3 + \beta_4 (d_{ik}^{RF})^3.$$  

The classification rates of the optimal $\pi_0$ values with respect to the cost functions are presented in Table 10. With the cost function $C_1(\pi_0, w_1 = 1, w_2 = 1)$, the best classifier is based on $M_{III}(V, V^{APC}, D)$. Comparing Table 10 with Tables 5, 6, 8, 9, and 7, we observe that the classifiers based on metric distance, volume, and APC in volumes usually perform better compared to the classifiers based on other models. Hence the model with volume, hippocampal volume loss over time, and metric distance is a better indicator for DAT compared to other models based on subsets of these variables.

The hippocampal volume change (together with the first 12 eigenfunctions) was used in [26] for logistic discrimination, but in this section, we use not the volume change but a function of it (namely annual percentage change) in logistic discrimination. Moreover, we also consider annual percentage change in metric distances in logistic discrimination.

5. Discussion and conclusions

In this study, we consider the use and analysis of metric distances between brain tissues that are computed from the vector fields generated by large deformation diffeomorphic metric mappings (LDDMM) with respect to an anatomical template. In particular, we use hippocampi in groups of subjects with and without Dementia of Alzheimer type (DAT) in its mild form (labeled as CDR0.5 and CDR0 patients, respectively) at baseline and follow-up as examples. The subjects in this paper have been previously analyzed using related but different tools. More specifically, the data was analyzed using the “greedy algorithm implementation of the diffeomorphic mapping and comparing displacement vector fields” [26]. [39] showed that this approach or the small deformation one used in many mapping methods do not generate true metric distances. Furthermore, in [26], as a single scalar measure, volumes were used for diagnosis group comparisons at baseline and follow-up and in [45] displacement momentum vector fields based on LDDMM were used for discrimination. At the same time [45] showed that statistical analysis of (momentum) vector fields generated by LDDMM was consistent with those generated by HDBM in [26]. But the metric distances computed from LDDMM has not hitherto been used in diagnosis group analysis. The metric distance gives a single number reflecting the global morphometry (i.e., the size and shape) while volume measurements mostly provide information on size. So metric distances provide additional morphometric information not conveyed by volume whereas momentum vector fields also provide local information on shape changes. Further, the morphometric information conveyed by the metric distance depends on the choice of the template, while the morphometric information conveyed by momentum vector fields is independent of the template chosen. That is, although the vector fields change when the template changes, the morphometric information they convey is the same. However, the vector field data is highly multidimensional and harder to interpret compared to the metric distances.

There have been other methods for hippocampus morphometry ([74,73,67,78,75–77] to name a few of the recent studies). Compared to these methods, the advantage of the current approach is that it enjoys all three of the following properties whereas the other methods only satisfy one or two of them. That is, the metric distance (1) provides a single global measure of shape and size, (2) can be used for group and discrimination analysis, and (3) indicates relative distance between subjects. For example, in [67], hippocampus was included as part of the temporal lobe and the tensor-based morphometry analysis was done using Jacobian maps created with a LDDMM type registration algorithm [79] but does not use the induced metric. Although the Jacobian maps calculated with LDDMM can be used for tensor-based morphometry analysis and can be compared with [67], it is beyond the scope of this work.

The metric distances may be used for cross-sectional comparisons of the morphometry of diagnostic groups. For example, in our example data set, metric distances did not detect any significant difference in morphometry at baseline (see Table 4), but follow-up metric distances for the right hippocampus in CDR0.5 (i.e., mildly demented) subjects are found to be significantly larger than those in CDR0 (i.e., non-demented) subjects (see Table 4). Wang et al. also analyzed the velocity vector fields for the baseline hippocampi of the same data set and found that the left hippocampus in the DAT group shows significant shape abnormality and the right hippocampus shows similar pattern of abnormality [45]. The reason for the metric failing to detect such abnormality in the baseline hippocampi is that metric distance is a compound and oversummarizing measure of global morphometry. The metric distances might also be used in longitudinal analysis in order to see how morphometry (with respect to the template) changes over time. For example, from baseline to follow-up, metric distances for CDR0.5 subjects significantly increase while those in CDR0 subjects do not (see Table 4). That is, the morphometry (shape and size) of hippocampus in CDR0.5 subjects changes significantly over time, but not in CDR0 subjects. Atrophy – over two years – might occur with aging, and this is captured by metric distances (see Table 4). However the increase in the metric distances in CDR0 subjects is not found to be statistically significant. Differences and changes (over time) in morphometry can also be used for diagnostic discrimination of subjects in non-demented or demented groups. Many discrimination techniques such as Fisher’s linear discriminant functions, support vector machines, and logistic discrimination can be applied to the metric distances, together with other qualitative variables. In this study we applied logistic discrimination based on metric distances, as logistic regression not only provides a means for classification, but also yields a probability estimate for having DAT. Furthermore, one can optimize the threshold probability for a particular cost function for the entire training data set, or by a cross-validation technique. The correct classification rate of the hippocampi was about 70% in our logistic regression analysis. In [45] PCA of the initial momentum of the same data set led to correct classification of 12 out of 18 (i.e., 67% of the) demented subjects and 22 out of 26 (i.e., 85% of the) control subjects. Metric distances can be used to distinguish AD from normal aging quantitatively; however, to be able to use it for
diagnostic purposes, the method should be improved to a greater extent.

The atrophy in the temporal lobe was also observed by [67], where hippocampus was included as a part of the temporal lobe (as opposed to our study where only hippocampus atrophy is analyzed). In [67], it has been shown that brain atrophy rates are greater in DAT and MCI (mild cognitively impaired) patients, compared to healthy subjects. Moreover, MCI subjects who converted to DAT exhibited faster atrophy rates than the ones who did not convert, and slower rates compared to DAT subjects. This is somewhat in

### Table 9
The correct classification rates ($P_{CCR}$), sensitivity ($P_{sen}$), and specificity ($P_{spe}$) percentages for the classification procedures based on models $M(C^{DC}_1) - M(D^{DC}_0)$ using APC in metric distances with optimum threshold values $\pi_1 = \pi_{opt}$ based on the cost function $C(\pi_{opt}, w_1, w_2)$ with $w_1 = w_2 = 1$ and $w_1 = 1, w_2 = 3$ (top); and with optimum threshold values $\pi_1 = \pi_{opt}$ based on the cost function $C(\pi_{opt}, \eta_1, \eta_2)$ with $\eta_1 = 3, \eta_2 = 7$ (bottom).

<table>
<thead>
<tr>
<th>$w_1 = w_2 = 1$</th>
<th>$w_1 = 1, w_2 = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

Using optimum $\pi_1$ based on cost function $C(\pi_{opt}, w_1, w_2)$ with

<table>
<thead>
<tr>
<th>$w_1 = w_2 = 1$</th>
<th>$w_1 = 1, w_2 = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

Using optimum $\pi_1$ based on cost function $C(\pi_{opt}, \eta_1, \eta_2)$ with

<table>
<thead>
<tr>
<th>$\eta_1 = \eta_2 = 5$</th>
<th>$\eta_1 = 3, \eta_2 = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

* The model with the best classification performance.

### Table 10
The correct classification rates ($P_{CCR}$), sensitivity ($P_{sen}$), and specificity ($P_{spe}$) percentages for the classification procedures based on models $M(D^{DC}_1) - M(D^{DC}_0)$ and $M(V, V^{DC}_1) - M(D^{DC}_0)$ using metric distance, and APC in hippocampal volumes with threshold probabilities with optimum threshold values $\pi_1 = \pi_{opt}$ based on the cost function $C(\pi_{opt}, w_1, w_2)$ with $w_1 = w_2 = 1$ and $w_1 = 1, w_2 = 3$ (top); and with optimum threshold values $\pi_1 = \pi_{opt}$ based on the cost function $C(\pi_{opt}, \eta_1, \eta_2)$ with $\eta_1 = 3, \eta_2 = 7$ (bottom).

<table>
<thead>
<tr>
<th>$w_1 = w_2 = 1$</th>
<th>$w_1 = 1, w_2 = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

Using optimum $\pi_1$ based on cost function $C(\pi_{opt}, w_1, w_2)$ with

<table>
<thead>
<tr>
<th>$w_1 = w_2 = 1$</th>
<th>$w_1 = 1, w_2 = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

Using optimum $\pi_1$ based on cost function $C(\pi_{opt}, \eta_1, \eta_2)$ with

<table>
<thead>
<tr>
<th>$\eta_1 = \eta_2 = 5$</th>
<th>$\eta_1 = 3, \eta_2 = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{opt}$</td>
<td>$\pi_{opt}$</td>
</tr>
<tr>
<td>$P_{CCR}$</td>
<td>$P_{CCR}$</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>$P_{sen}$</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>$P_{spe}$</td>
</tr>
</tbody>
</table>

* The model with the best classification performance.
agreement with our findings that the rate of change in morphometry is larger in DAT subjects compared to healthy ones. However, volume (as a measure of tissue size) is not analyzed in this article. Furthermore, compared to the vector field changes, Jacobian changes are used in the analysis of morphometric changes.

We perform a principal component analysis on metric distances and hippocampus, brain, and intracranial volumes. Considering the variable loadings, we conclude that volumes are mostly measures of size and partly related to shape, while the metric distance is mostly a measure of size and partly related to size.

We also compare the cross-sectional, longitudinal, and discrimination results of metric distances with those of volumes. We observe that cross-sectional and longitudinal analysis give similar results, although metric distances increase and volumes decrease by time. The metric distance, being an extremely condensed summary measure gives very similar results as the hippocampal volume. Furthermore, the differential volume and distance changes are measured by annual percentage rate of change (APC) for the two year period in the study. Similar to the results of [66], we found that APC in volumes may be a good indicator for early stage of DAT. However, APC in metric distances do not provide a good performance in classification of CDR0 versus CDR0.5 hippocampi. The best classifiers in our logistic discrimination procedure are the ones which use hippocampal volume, metric distance, and annual percentage change in volumes (see Table 10). Notice that the best performer models (marked with *) have correct classification rates of 91.1%, sensitivity rate of 94%, and specificity rate of 88%. On the other hand, in [26], the logistic discrimination with the first 12 eigenfunctions based on SVD yields a correct classification rate of 84.0%, sensitivity rate of 83.3%, and specificity rate of 84.6%; and the logistic discrimination with the first 12 eigenfunctions together with hippocampal volume changes yields a correct classification rate of 86.0%, sensitivity rate of 72.2%, and specificity rate of 92.6%. Therefore, in classification of healthy versus diseased hippocampi, our model (with volume, metric distance, and annual percentage change in volume) seems to perform better compared to the classification models of [26] in the classification rates. Hence these measures may constitute a reliable biomarker when used together.

The clinical implications of the findings are deferred to [68]. We have presented detailed statistical analysis of metric distances computed with LDDMM and show that this is potentially a powerful tool in detecting morphometric changes between diagnosis groups or changes in morphology over time. Metric distances depend on the choice of template anatomy used. However, in this article we do not address the issue of template selection for optimal differentiation between hippocampus morphology. As a compound but brief measure of morphometry, metric distances can thus serve as a first step to identify the morphometric differences, and can be used as a pointer to which direction a clinician or data analyst could go. The metric distance results agree with the volume comparisons of [26], hence volume (i.e., scale) might be highly dominating the morphometric changes in the hippocampi. In other words, the significant volume reduction in left and right hippocampi might dominate the change in shape, when morphometry is measured by metric distances. To remove the size influence so as to measure the shapes only, one can perform scaling on the hippocampi and then apply LDDMM to normalize the size differences.

Acknowledgments

We would like to thank anonymous referees, whose constructive comments and suggestions greatly improved the presentation and flow of the paper. This research was supported by: Pacific Alzheimer Research Foundation, Michael Smith Foundation for Health Research, Canadian National Science and Engineering Research Council (NSERC) and NIH grants (P50 AG05681, P01 AG03991, P41 RR15241).

References


Lei Wang received his PhD from Harvard University in 1995. Currently he is an assistant professor in the Department of Psychiatry and Behavioral Sciences and Department of Radiology at Feinberg School of Medicine, Northwestern University in Chicago. His research interests include Alzheimer’s disease, aging, schizophrenia, bioinformatics, psychiatry, radiology, and bioengineering.

John C. Morris received his BA from Ohio Wesleyan University and his MD from the University of Rochester School of Medicine and Dentistry in Rochester, NY, in 1974. Currently he is the Harvey A. and Dorismae Hacker Friedman distinguished professor of neurology, professor of pathology and immunology, professor of physical therapy, and professor of occupational therapy at Washington University. He also is the director and principal investigator of the Alzheimer’s Disease Research Center. The focus of Dr. Morris’ research and practice is Alzheimer’s disease and other neurological disorders associated with aging. Specific research interests include detecting preclinical Alzheimer’s disease, improving the diagnosis of early-stage Alzheimer’s disease, evaluating new drugs for the treatment of dementia, and establishing phenotypes for inherited forms of Alzheimer’s disease and other dementias.

John G. Csernansky received his MD degree from New York University School of Medicine in 1979, and completed his training in psychiatry and psychopharmacology at Stanford University in 1985. Currently, he serves as the Gilman Professor and Chairman of the Department of Psychiatry and Behavioral Sciences at the Feinberg School of Medicine at Northwestern University in Chicago, Illinois. His research in the areas of clinical interest includes schizophrenia, Alzheimer’s disease, personality disorders, neurodegenerative disorders and in general includes the investigation of relationships between neurobiological measures, specific dimensions of cognition and psychopathology.

Michael I. Miller received his BS at Electrical Engineering from State University of New York at Stony Brook in 1976, an MS at Electrical Engineering from Johns Hopkins University in 1979, and PhD at Biomedical Engineering from Johns Hopkins University in 1984. Currently he is the Herschel and Ruth Seder Professor of Biomedical Engineering, professor of electrical and computer engineering, and the director of Center for Imaging Science. His research interests include image understanding and computer vision; medical imaging and computational anatomy and computational neuroscience.

J. Tilak Ratnanather received his PhD in mathematics at the University of Oxford. His postdoctoral research at City University and Johns Hopkins University School of Medicine focused on thermal boundary layer separation and mathematical models of cochlear physiology and micromechanics. Since 1998 he has been on the staff and now faculty at the Center for Imaging Science where he works on developing and applying mathematical methods in shape analysis of cortical and sub-cortical structures in a variety of neuro-degeneration and neuro-developmental disorders.