Topology based Kernels with Application to Inference Problems in Alzheimer’s disease

Deepti Pachauri,
Department of Computer Sciences, UW-Madison

Chris Hinrichs,
Department of Computer Sciences, UW-Madison

Moo K. Chung,
Department of Biostatistics and Med. Informatics at UW-Madison

Sterling C. Johnson, and
Wisconsin ADRC (Department of Medicine) at UW-Madison and William S. Middleton VA Hospital

Vikas Singh
Department of Biostatistics and Med. Informatics, Department of Computer Sciences, and
Wisconsin ADRC at UW-Madison

*for the Alzheimer’s Disease Neuroimaging Initiative*

Deepti Pachauri: pachauri@cs.wisc.edu; Chris Hinrichs: hinrichs@cs.wisc.edu; Moo K. Chung: mkchung@wisc.edu;
Sterling C. Johnson: scj@medicine.wisc.edu; Vikas Singh: vsingh@cs.wisc.edu

Abstract

Alzheimer’s disease (AD) research has recently witnessed a great deal of activity focused on developing new statistical learning tools for automated inference using imaging data. The workhorse for many of these techniques is the Support Vector Machine (SVM) framework (or more generally kernel based methods). Most of these require, as a first step, specification of a kernel matrix $K$ between input examples (i.e., images). The inner product between images $I_i$ and $I_j$ in a feature space can generally be written in closed form, and so it is convenient to treat $K$ as “given”. However, in certain neuroimaging applications such an assumption becomes problematic. As an example, it is rather challenging to provide a scalar measure of similarity between two instances of highly attributed data such as cortical thickness measures on cortical surfaces. Note that cortical thickness is known to be discriminative for neurological disorders, so leveraging such information in an inference framework, especially within a multi-modal method, is potentially advantageous. But despite being clinically meaningful, relatively few works have successfully exploited this measure for classification or regression. Motivated by these applications, our paper presents novel techniques to compute similarity matrices for such topologically-based attributed data. Our ideas leverage recent developments to characterize signals (e.g., cortical thickness) motivated by the persistence of their topological features, leading to a scheme for simple constructions of kernel matrices. As a proof of principle, on a dataset of 356 subjects from the ADNI study, we report good performance on several statistical inference tasks without any feature selection, dimensionality reduction, or parameter tuning.
I. Introduction

Alzheimer’s disease (AD) is a neurodegenerative disorder that leads to progressive loss of memory and cognitive function. Early clinical diagnosis is challenging because AD-specific changes begin years before the patient becomes symptomatic [1], [2], [3], [4], [5]. Therefore, a major focus is to identify such changes at the earliest possible stage by leveraging neuroimaging data [6]. One specific interest is to go beyond group analysis (i.e., between clinically different groups), rather to “learn” patterns characteristic of neurodegeneration using machine learning (ML) methods. Of course, AD is just one potential application, and image-based machine learning methods are applicable to a variety of neuroimaging problems.

Statistical machine learning methods provide means for learning a hypothesis (or concept) from a set of training examples. In the context of a specific disorder (e.g., AD), the training data is given as a set comprised of both diseased and control subjects, and our objective is to learn a pattern in such examples to help predict the target variables for new “test” cases. Some work in the last few years has focused on making use of a popular machine learning algorithm called Support Vector Machine (SVM) learning for inference problems, such as classification, regression etc. A main focus of recent efforts has been directed towards classification of AD and control subjects in the hope that this will lead to earlier predictions of which subjects will go on to develop AD. Beyond the classification domain, however, there are other significant scientific questions which may be addressed via kernel methods, e.g., analyzing subtle interactions between biological or psychological measures and brain topology. We provide a brief review of these strategies next to highlight their key advantages and limitations, especially in the context of the data types (and the inference problems) of interest in this paper. The review will not only help establish the discriminative power of imaging data (esp. cortical thickness), but highlight the growing consensus towards the design of automated tools, so as to use other markers of histological changes (plaques, tangles, etc.) in a disease’s presymptomatic phases by adapting extensions of ML algorithms.

A. Related Work

The need to represent, manipulate, and analyze complex brain data has led to a set of novel techniques for brain surface parameterization. These algorithms provide a smooth functional representation of the complex inherent geometries of brain surfaces [7], [8], [9], and have successfully been applied to discover disease-specific regions in group analysis [10]. On the other hand, a number of recent efforts have focused on utilizing domain knowledge. For instance, in [11], the authors exploited the fact that the hippocampal volume was atrophic in AD and reported 82%–95% classification accuracy on a data set of 39 subjects (19 AD, 20 controls). Numerous MR imaging-driven volumetric studies have used Voxel Based Morphometry (VBM) [12] to perform disease specific statistical analysis using clinically relevant measures at the voxel level (such as gray matter density [13]), and show differences in regional gray matter loss between the two groups [14]. However, [15], [16], [17] have independently shown that classification models that depend on voxel-by-voxel evaluation of the neurological data require extensive feature reduction either by recursive elimination [18] or by manipulating the weights learned by a linear SVM [16]. Klöppel and colleagues [16] achieved 95% accuracy on a data set of 90 subjects (33 AD, 57 controls). Likewise, classification performance of methods based on PCA [19], [20] show a dependence on

Index Terms

Cortical thickness based kernels; topological persistence; Alzheimer’s disease; ADNI
selection criteria of principal components of the dataset, while [21] proposes that improvements in accuracy can be obtained via a spatial regularization (Markovian interaction between voxels) and reported an increased classification accuracy of 82% on a subset of 183 subjects from ADNI. While feature selection (or dimensionality reduction) is extensively used, it involves a careful selection of parameters to preserve the important components of the signal. Also, the performance must not be too dependent on the sample size, and the number of features must remain stable across datasets [22]. Nonetheless, the above papers demonstrate that good classification accuracy on certain types of neuroimaging data can be obtained via novel pre-processing or by incorporating additional domain specific modifications in the learning algorithm. Next, we make the case that the above techniques are not easily extensible to a class of (clinically interesting) neuroimaging data for many inference tasks, and new methods are needed to exploit the discriminative components of the signal.

Recent evidence suggests that neurological measures defined on cortical surfaces, such as cortical thickness values [23], [24], [25], [26], [27] are a clinically meaningful measure of the underlying brain morphology and a reliable measure of gray matter atrophy [28], [29], but are not yet fully utilized for automated inference in the context of AD. To address this problem, we focus on methods for extraction and precise characterization of important topological features for such data – with cortical thickness values as an example, we design new methods for “kernelizing” signals defined on cortical surfaces. Using characterizations inspired from the “persistence” of topological surface features, we present a framework for construction of affinity (or weight) matrices between subjects’ distributions of cortical thickness (or other measures 1), each defined on surfaces. This permits using such data systematically within any off-the-shelf statistical toolbox, without manual parameter or feature selection. The construction of similarity measures based on a topological signal such as cortical thickness allows the inclusion of topological data along with other biomarkers in predictive models. The inclusion of such highly independent covariates offers the potential to aid machine learning methods that operate with ROI measures, gray matter maps, and so on, by providing additional information which is not detected by other feature extraction methods. Note that our goal is not to argue that cortical thickness is better than existing features, rather, the machinery developed here presents methods to construct representations for such data which may then be used in conjunction with other features of choice. Indeed, since cortical thickness is a clinical measure, its relevance will be different in different application scenarios. Finally, the proposed algorithm achieves a general purpose characterization of any topology-based signal in a mathematically elegant framework – which is in fact the primary focus of the proposed method.

The contributions of this paper are: (i) We give topology-based kernel construction algorithm for measures defined on cortical surfaces. (ii) On a dataset of AD and control subjects (ADNI), we investigate the effectiveness of proposed characterization in group level studies, standard regression, and classification. This avoids the difficulties of extensive preprocessing and regularization, as when using standard ML tools. In an additional set of experiments with standard ROI-based features, we evaluate the utility of proposed topology based features. (iii) We provide an implementation for performing inference (classification, regression, linear models, and so on) on complex data such as signal defined on cortical surfaces. The goal of this work is neither to assert that cortical thickness is the best discriminator of AD, nor to claim that the proposed method is the best mechanism for classification purposes. Rather, it is a general purpose non-parametric method for

---

1The complete analysis in the paper refer to cortical thickness values. We will be using terms “signal”, “brain measure” and “cortical thickness values” interchangeably because the proposed method is easily extensible to any brain measure defined on cortical surfaces.
constructing similarity measures on topological signals which can be included in multi-modal or multi-variate models to boost the signal of interest.

II. Main ideas

In the following section, we draw an outline for the motivation behind our method in the context of AD-related statistical inference. We will present the complete formulation of topology based kernel construction using cortical thickness signal in the next section.

In the context of Alzheimer’s disease as well as other disorders [30], [31], [32], numerous findings confirm fundamental differences in the patterns of cortical volume loss, and regard cortical atrophy a useful biomarker for AD [10], [33]. Cortical thinning leads to localized changes in the spatial distribution of gray matter, and hence a geometric realization of topological measures on the cortex can induce separability between diseased and healthy brains. The goal is to use such clinical signals to derive similarity measures. Because cortical thickness is a highly attributed marker, most of the existing models have used this anatomically relevant aspect of disease progression in a generative setting by assuming co-registered surfaces and by performing statistical tests to evaluate the discriminative power of each feature, where the feature vector’s dimensionality equals the number of cortical surface vertices, and the thickness values at the vertices give the magnitude of each corresponding entry. One may then calculate inter-subject distances (or similarities) in terms of geometric distances based on top discriminative features (or inner products), which may then be fed into a standard SVM procedure. Notice the two major difficulties with this approach: first, we must account for the mismatch between the training set size and the dimensionality of the distribution via feature reduction [22], [19] or introduction of bias [21]. Second, during the reconstruction step of cortical thickness using automated software tools, point-wise correspondence of mesh topology among the training subjects may be unavailable. This will lead to different number of vertices at different coordinates for different subjects. One approach to this problem is to try mapping cortical thickness onto a sphere with a fixed number of vertices, re-sample and interpolate the cortical thickness measure for each subject so as to allow a direct point-wise comparison. However, the procedure not only attenuates vertex-wise signal but also ignores the higher order interactions between subsets of vertices that vary between the two groups. Also neurodegeneration is an exclusive event, and exact locations (coordinates of affected vertices) also vary among the population, obscuring the statistical concept under study. More importantly, in the context of cortical thickness, vertex-wise thickness values may be less relevant – in some settings separability between classes likely comes from variation between topological features (comprised of more than one vertex). In other words, subtle losses of gray matter may affect the shape or topology of cortical surfaces before they significantly affect the thickness measures. A naive alternative is to look at all possible groups of vertices, and evaluate the significance of their variation, which is clearly intractable. Our approach below seeks to characterize brain images by deriving a representation of “groups” of vertices on each individual cortex. Briefly, we determine whether a localized region in the brain exhibits any gray matter atrophy via construction of a simplex on critical points – which provide information on the global topology.

Our method is based on topological persistence of brain measurements defined on cortical surfaces. Within the framework of signal filtration, if one considers the cortex as a complex then the spatial properties of cortical thickness can be represented as the history of a growing complex using notions of “birth and death” of homological classes. It is reasonable

---

1Note that this means that in the case of AD vs. control classification, such an advantage may be lost, since cortical thickness has been shown to be discriminative by itself.
to expect that inferences based on such topological changes which occur during the growth either as critical points or noise as a function of their lifetime, will capture the differences between clinically different groups well on a global level. Our algorithm seeks to characterize such changes to derive a precise representation of the topological features on the signal. Once such a representation is obtained, we can easily construct similarity measures (kernels) and leverage emerging ML tools (e.g., MKL) for inclusion in statistical inferences.

### III. Algorithm

#### Summary

Let us denote the set of cortical data in the training set as $X = \{X_1, X_2, \ldots, X_n\}$ with known target variables $y = \{y_1, y_2, \ldots, y_n\}$, where $y_j \in \{+1, -1\}$ (or $y_j \in \mathbb{R}$) in a classification (or regression) setting. $X$ comprises the cortical data of both AD and control groups. The original signal defined on the cortical surface is noisy and it is necessary to estimate the unknown signal to increase signal-to-noise ratio and smoothness of the signal for subsequent analysis. This is achieved via image smoothing over $\mathbb{S}$. Making use of the mean signal is not ideal for constructing similarity matrices which promote separability, due to signal attenuation. Such a process will make it rather difficult to identify subtle, but clinically relevant variations. Further, substantial overlap in the class distributions confounds inference. In contrast, our proposed method is motivated by the topological characterization of the signal which makes no assumption on point-wise correspondences of cortical measure on a naturally formed spherical atlas. Hence, no assumption is made regarding a specific pre-processing tool for extraction, allowing the proposed framework to handle disparities among spherical atlases of individual subjects (generated by some pre-processing tool of choice). As a result, the collection of critical points on the mean signal is adequate to characterize the topology of the signal. By pairing the critical points in a nonlinear fashion, we construct scatter plot for each image, which encodes the topological properties of the input signal. The concentration maps of such scatter plots are then used to learn the disease patterns by employing off-the-shelf ML toolbox. We now present details of the algorithm in the following subsections.

#### A. Heat Kernel Smoothing

Existing methods model the cortical measurement $f(x)$ as signal plus noise: $f(x) = \mu(x) + \epsilon(x)$, where $\mu$ is the signal, and $\epsilon$ is noise. In the following analysis, $f(x)$ is the cortical thickness mapped onto a spherical surface, i.e., $x \in \mathbb{S}$ as shown in Fig. 1 (right).

We used heat kernel smoothing to estimate the signal [7], [34]. The heat kernel is essentially Green’s function of the isotropic diffusion equation [34]. The solution of which with the following initial conditions ($f(t, \sigma = 0) = \delta$) provides the heat kernel $K_\sigma$ as

$$K_\sigma(t, s) = \sum_{l=0}^{\infty} e^{-\lambda_l \sigma} \sum_{m=-l}^{l} Y_{ln}(t)\hat{Y}_{ln}(s) \quad (1)$$

where $-\lambda_l = l(l+1)$ is the eigenvalue and $Y_{ln}$ are eigenfunctions of the operator known as spherical harmonics. Spherical diffusion of a given functional measurement $f \in L^2(\mathbb{S})$ i.e., a smooth estimate of $f$ denoted as $\hat{f}$ is written as a convolution: $\hat{f} = K_\sigma * f$. The finite estimate of the continuous convolution is called the Weighted Fourier Series (WFS).
where $\hat{f}_{lm} = \langle f, Y_{lm} \rangle$ is the Fourier coefficient. This is a more flexible spectral approach that explicitly represents the solution to the diffusion equation analytically.

In principle, series approximation of smoother cortical signal improves by increasing the number of coefficients in the summation. The optimal choice of the degree is useful to save computational time. Fig. 2 (top row) shows the finite expansion approximation of the cortical thickness shown in Fig. 1 as the degree increases from $k = 5$ to $k = 70$ (bandwidth ($\sigma$) = 0). The spatial smoothness of the thickness is controlled by $\sigma$. Fig. 2 (bottom row) shows the thickness data as $\sigma$ varies in the range $[10^{-2}, 10^{-5}]$. The goal is to estimate the signal using WFS to suppress noise. An optimal selection of $\sigma$ is required to avoid overfitting and not to lose the discriminative features during smoothing step. We used $\sigma = 10^{-4}$ and corresponding degree $k = 50$ in our experiments based on [7].

B. Scatter plots of Persistence

We will characterize the signal in terms of its critical points and the sublevel sets. The sublevel sets of a signal $f$ are defined as $R(y) = f^{-1}(-\infty, y]$ where $f(x)$ is the signal (i.e., the corresponding cortical thickness measure), and $f^{-1}(y)$ is its inverse. In other words, a sublevel set is the set of all points $x$ for which $f(x) \leq y$. Observe that critical points of the signal (i.e., local minima and maxima) provide information about local variations in signal. But we are interested in points which preserve information about the connected regions of its sublevel sets. We can define the time-point (thickness value) at which a critical point will be “visited” in a filtration of the signal [35]. As we increase $y$, the critical points of the signal are visited (in order) and homology of $f(x) \leq y$ changes (with $y$) i.e., signal value $f(x)$ is critical if the rank of homology groups of sublevel set $R(f(x))$ change by visiting $f(x)$, i.e., if a homology class is born (or dies) by visiting $f(x)$.

Clearly, the critical points are markers of the creation and deletion of connected components in the sublevel sets, with increasing values of the filtration. More importantly, unique pairings of these critical points are sufficient to identify each topological feature in the signal. Our pairing rule is defined using persistence of critical points [36], [37] and is formulated as follows:

Consider an ordered set $C$ of all critical points of $f(x)$ (i.e., cortical thickness), $C = \{c_1 \leq c_2 \leq \ldots \leq c_n\}$. This defines a sequence of connected components and a corresponding sequence of homology groups. Our aim is to store the history of homological changes within this sequence i.e., a history of homology classes. A class is born at $c_j$ if it did not exist in sublevel sets $R(y)$ such that $y \in C_j$ where $C_j = \{c_j : \forall c_j \geq c_j, c_j, c \in C\}$. A class dies entering $c_h$ if it existed in sublevel sets $R(y)$ such that $y \in C_h$ but did not exist in sublevel sets with filtration value in subset $C_h = \{c_j : \forall c_j \geq c_h, c_j, c_h \in C\}$. Now we can represent the “life duration” of a class born at $c_j$ and dies at $c_h$ by pairing the critical points $(c_j, c_h)$.

Furthermore, we can present the history of each class by drawing these pairs of critical points in two-dimensional plane “scatter plot”. The construction of a scatter plot for an arbitrary one-dimensional function $f(x)$, $x \in \mathbb{R}$ is illustrated in Fig. 3.

To facilitate the construction of scatter plots using cortical thickness for each subject, we first calculated critical points of the cortical thickness. Critical points in this setting are defined as those points where the cortical thickness is maximum/minimum relative to its immediate neighbors. We used Delaunay triangulation [38] to generate a simplicial complex of critical points. The signal value was used to define the “filtration” on the complex. We
used publicly available software [39] to calculate the persistent homology in degree-0 for each subject. The procedure essentially provides a unique pairing of critical points which preserve the information of connected regions as discussed above. That is, preserving only the paired critical points (birth-death pairs) and the corresponding filtration step at which they were visited i.e., thickness values. This procedure yields a collection of topological feature descriptors (i.e., connected regions of a sublevel set) for each subject which can be represented as a scatter plot (also known as a plot of Betti intervals). In the scatter plot representation, each feature is described in terms of its birth and death “time”, i.e., the filtration values of the paired critical points. Now, by comparing the scatter plots of such paired values, we can establish the desired similarity measure between subjects.

We construct a concentration map from the scatter plot by using kernel density estimation (KDE); we estimate the distribution of pairs on a uniform grid. Vectorization of the concentration map provides a probability distribution function (PDF) for each subject. This is PDF of the scatter plot, and does not directly tie to the high and low thickness values. By choice of a suitable KDE similarity measure, we can now compute similarity inter-subject measures, that will give a positive semi-definite (PSD) kernel matrix; where entries are the similarity between these PDFs. If the topology of cortical thickness is discriminative then many of these cortical topology measures will show a variation (between AD and controls) resulting in measurable group differences in the clinical populations in terms of their critical point pairings. Observe that with the similarity measure and kernels constructed above, the application of kernel regression or classification (also, group analysis) on this data is quite simple. In Fig. 4 we have demonstrated the various steps involved in the algorithm.

IV. Experimental setup

Our objective in this section is to evaluate the effectiveness of the chosen topological representation. As an illustrative example, we focus on inference problems in AD such as classification, regression and group studies on clinically different groups. The purpose here is not to achieve the best classification accuracy for AD (or that cortical thickness is the most discriminative feature) rather to show how topologically-based data can be systematically used within image-based statistical analysis. We performed experiments using cortical thickness signal defined on cortical surfaces for each subject employing our algorithm discussed in III. The initial processing of T1 weighted MR images was performed using FreeSurfer. To learn discriminative topological characteristics we used a set of 356 subjects (160 AD, 196 controls) in our evaluations. We made use of ground truth diagnosis of each subject (i.e., class labels), based on given clinical evaluation of cognitive status for each subject. A more detailed description of the data set, preliminary processing and algorithm implementation are covered in the following subsections. Later, in section V, we demonstrate the use of derived features in various inference settings.

A. Data set

We evaluated the accuracy of our method by performing experiments on data collected as part of the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (www.loni.ucla.edu/ADNI). A summary of demographic and neuropsychological details of subjects used in our study is presented in Table I.

B. Preliminary data processing

The MR images in our data set were processed using FreeSurfer [41] to calculate the cortical and subcortical anatomy. To enable the reconstruction of brains cortical surface from structural MRI data, freely available software FreeSurfer (v1.133.2.57) was used on a 64-bit Linux workstation. FreeSurfer models the boundary between white matter and cortical gray
matter, tessellates the resulting surfaces, and calculates distances between the surfaces (i.e., thickness values) at each vertex. This procedure gives an cortical thickness measure at each point on the cortex for each subject’s cortical surface (in fact, for each pair of left and right hemispheres) in our dataset\(^2\). For improved visualization, a cortical surface-based spherical atlas has been defined based on average folding patterns (see Fig. 1). Surfaces from individuals were aligned with this atlas with a nonlinear registration algorithm. The registration is based on aligning the cortical folding patterns and therefore directly aligns the anatomy instead of image intensities. The spherical atlas forms a coordinate system in which point-wise correspondence between subjects can be achieved by interpolation, though such correspondence is not required for our algorithm.

C. Implementation

Mean signal on the cortical surface was obtained via image smoothing, as discussed in III(A). We calculated critical points and scatter plots on the mean signal for each subject Fig. 5(a), as discussed in III(B). To derive a similarity measure on scatter plots, we used kernel density estimation to generate the concentration map of a scatter plot on a square grid \([0, 5]^2\) with 50\(^2\) pixels for each cortical surface Fig. 5(b). Once the concentration map for each cortical surface was obtained, vectorized concentration maps (PDFs) Fig. 5(c) were used to construct the kernels. We used these kernels to perform the following experiments:

i. 2-sample \(t\)-tests on the concentration maps (PDFs) to locate statistically relevant pixels in the concentration map.

ii. Kernel regression to model important cognitive markers in a 10-fold cross-validation setting [42] for the two clinical groups. (iii) 10-fold cross-validation classification to estimate predictive accuracy on unseen test examples using the kernel SVM implementation provided in the Shogun toolbox [42]. The entire process described above was carried out separately for the right and left hemispheres of each subject’s brain, giving two trained SVM classifiers. While sophisticated methods to combine different types of kernels exist, our final prediction label for a test example was derived by averaging the two classifier outputs – i.e., winner (classifier with a higher confidence) takes all. An additional set of classification experiments to assess the predictive accuracy of ROI-based features supplemented with topological persistence kernels is also presented. (iv) Localizing the brain regions which were involved in homological changes in the two groups.

V. Experimental results

In order to show the effectiveness of topological features, we have performed (A) group analysis, (B) regression, and (C) classification. Here, we present an analysis of the performance of our algorithm on the data described above.

A. \(t\)-statistics

In order to test the hypothesis that the proposed topology based features do indeed extract meaningful about AD-related atrophy, we performed 2-sample \(t\)-tests on each pixel in the concentration map. Fig. 6 shows the differences in means for each pixel above (AD - control, for left and right hemispheres,) and \(p\)-values in negative log scale below, with statistically significant pixels in correspondingly “warmer” colors. Several notable properties can be seen from the figure. First, we see that there is a significant shift in the AD

\(^2\)AD brain scans have severe atrophy due to which “auto-Talairach” step failed. Therefore during automated cortical reconstruction process using recon-all directive, we disabled Talairach transformation via –notal-check for all subjects.
group towards the origin, corresponding to the global thinning seen in AD. Second, we can see from the distribution of the $p$-values that there are more subtle effects in regions off of the $y = x$ line, which suggest that there are higher-order topological effects. Finally, we can see that there are regions of non-zero density below the $y = x$ line despite there not being any min-max pairs in these regions – this is due to the kernel density estimation step, which interpolates the density function to regions where there are no min-max pairs, and which is agnostic of the process by which the points are extracted.

B. Regression

We also used our topological representation to predict the values of various relevant cognitive measures in a 10-fold cross-validation setting using a kernel regression model. This way, the model does not have access to the quantities it is trying to estimate until after training. We then performed 2-sample $t$-tests between the outputs to show that the model is indeed detecting separable group differences. We found that predicted values of 20 neuropsychological scores showed highly significant group-wise differences between AD and control subjects. The $p$-values of two-tailed $t$-test were $\approx 0$ for these neuropsychological scores. In other words, the estimated values of individual cognitive scores were discriminative for clinically different groups. This also helps explain the fact that gray matter atrophy, and thus topological features lead to cognitive decay. Table II presents the complete set of significant results.

C. Classification

In addition we calculated predictive accuracy, ROC curves, and area under ROC curves for cross-validated classification experiments. These results are shown in Table III and Fig. 7, and summarized below. The combined classifier had an accuracy of 75%. Left hemisphere- and right hemisphere-derived classifiers had accuracies of 73.31% and 73.03% respectively. The higher accuracy of the combined classifier compared to that of either of the single-hemisphere-derived classifiers suggests that it affords greater separability of the clinical groups when both hemispheres are used. The AUC measures were 0.7955 for the left hemisphere derived classifier, 0.7973 for the right hemisphere derived classifier, and 0.8063 for the combined classifier.

1) ROI—In order to quantify the significance of proposed topology based features, we investigated the following hypothesis: Do topology persistence based kernels add to the discriminative power of ROI-base kernels? If so, we can conclude that the proposed representation scheme provides information on certain aspects of the concept not captured by ROI features on their own.

ROI-data from ADNI consists of volumetric numerical summaries of hippocampus and temporal lobes. In a set of experiments, we studied the predictive accuracy for AD vs CN of ROI-base features exclusively. In the second step, we combined the topological & ROI-based kernels and tested the discriminative power of the unified kernel on exactly same 10-fold setup used in ROI-based experiments. Table IV shows the different combinations and their predictive accuracies. Except one combination (left middle temporal), all combinations show non-negligible improvements in the mean accuracy. We performed a paired $t$-test on the outcome. We found that while on all cases the mean accuracy improved, for the left & right hippocampal ROIs as well as right middle temporal lobes, the null-hypothesis can be rejected at $\alpha = 0.05$. For the two other features, the $p$-values were also small. This experiment, provides reasonable evidence that topological persistence based kernels add relevant information in the unified kernel for statistical analysis tasks.
D. Correlation between algorithm outcomes and cognitive biomarkers

From both a therapeutic and a research standpoint, it is important to identify signs of a disease at its predementia stage. Current diagnostic criteria for AD are based on neuropsychological scores such as: mini-mental status examination (MMSE), and neuropsychological battery scores such as logical memory and delayed recall. Other neuropsychological scores (Category fluency, Trail making, Rey auditory, Boston Naming etc.) are not used in diagnosis, however, as measures of cognitive status we nevertheless expect such scores to be highly correlated with the diagnosis (i.e., ground truth), and regard them as “pseudo-ground truth” due to this confounding influence. Recent studies have suggested that in early stages of AD, cognitive impairment does not correlate with brain structure delineation, which explains the weakness of cognitive scores particularly with early AD onset identification. Fig. 8, illustrates statistical correlation of many cognitive scores with the classification confidence output by an SVM trained on the proposed imaging based topological features. We found that most cognitive scores are in agreement with the classification confidence for both left and right hemisphere. This suggests that indeed imaging based markers are capable of predicting the status of individual subjects accurately so as to identify individuals likely to progress to AD.

E. Localizing persistence

Finally, we used spatial information of min-max pairs to locate the brain regions persistently involved in topological changes in each subject. These are representative locations on the brain where homology classes were born (and died) during the growth of individual cortex. Fig. 9 (top) shows the differences in means of min-max locations for the clinical groups (AD - control, for left and right hemispheres). We noticed significant differences in the patterns of topologically active brain regions between the two groups corresponding to the global thinning in AD. Fig. 9 (bottom) shows p-values in negative log scale of 2-sample t-tests on each vertex. Statistically significant regions are shown in “warmer” colors. Noticeably, localized regions (neighboring vertices) corresponds to the cortical thinning in specific brain regions.

VI. Conclusions

In this paper we have proposed a new method for constructing kernels using cortical thickness measures defined on brain surfaces. This type of feature extraction will allow the inclusion of highly attributed brain signals into statistical inference machinery. We have demonstrated the value of these features in the context of mild AD via kernel based analysis methods. Our method offers a discriminative, yet compact representation of cortical surfaces, using persistence of the topological features induced by cortical thickness signal. The method is self-governing with respect to any point-wise correspondence of the mesh topology among subjects and any kind of feature reduction. The underlying concept of our method enables easy computation of meaningful similarity measures in a non-generative manner for use within machine learning methods, not offered by most existing methods. Due to this reason, our experiments show possible and scientifically important inferences that can be derived using proposed features, using AD as an illustrative example. Our experimental results demonstrate the strength of topological features in capturing subtle anatomical variations present in complex data. Given that cortical thickness measures are suspected to be relevant for neurological disorders. In its present form, the paper outlines mechanisms to exploit such data, if available. We believe that these methods will add to the diagnostic accuracy obtainable by MR and/or FDG-PET images alone [21], or within multi-modal analysis frameworks proposed recently [43], [44]. Our implementation will be provided at http://pages.cs.wisc.edu/~pachauri/minmax.
Acknowledgments

This work was supported in part by NIH grants R21-AG034315 (Singh), R01-AG021155 (Johnson) and P50-AG033514 (Asthana). Funding for this research was also provided by the University of Wisconsin–Madison UW ICTR through an NIH Clinical and Translational Science Award (CTSA) 1 UL1 RR025011, a Merit Review Grant from the Department of Veterans Affairs, and Wisconsin Comprehensive Memory Program. The authors also would like to acknowledge the facilities and resources at the William S. Middleton Memorial Veterans Hospital. In particular, we would like to thank Dr. Guofan Xu and Jennifer Oh for their help in data pre-processing.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI; Principal Investigator; Michael Weiner; NIH grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions of the following: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer’s Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging, with participation from the U.S. Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

References


Appendix A

With an eye towards making our paper as self-contained as possible, in the following we give a viable illustration of scatter plot construction for an arbitrary one-dimensional function $f(x)$, $x \in \mathbb{R}^1$. For more information pertaining to homological classes and persistence, see [36], [45], [46], [40].

Fig. 10 shows a one-dimensional signal $f(x)$. This function has four extremum points, two local minima $[c_1, c_2]$ and two local maxima $[c_3, c_4]$. Sublevel set of $f(x)$ with $R(y) = f^{-1}(c_1)$, is shown as collection of data points in solid red color under blue line in Fig. 11(a).

Similarly, other sublevel sets corresponding to other extremum points $c_2$, $c_3$ and $c_4$ are shown in Fig. 11(b), (c), and (d) respectively.

The pairing scheme of critical points in one-dimensional case, can be understood as follows: the rank of homology groups of all sublevel sets $R(y)$ with $y \preceq c_2$ is one: only $H_1$ (one-dimensional hole) is present. At time-point $c_2$ ($f(x) = c_2$), the rank of homology groups for sublevel set $R(y = c_2)$ becomes two with the “birth” of another class $H_2$ by entering $c_2$. For sublevel sets with $R(y)$ such that $c_2 \leq y \preceq c_3$, the rank remained constant. Entering $c_3$, the class born at $c_2$ ($H_2$) “died” and the rank of homology groups becomes one for sublevel set $R(y = c_3)$. Therefore, the persistence (history) of class $H_2$ is preserved in time-points $(c_2, c_3)$.
Similarly, pair \((c_1, c_4)\) represents the persistence of homology class \(H_1\). Fig. 12 demonstrate the pairing scheme for topological persistence. For completeness, Fig. 15 is shown with the one-dimensional signal \(f(x)\), Fig. 15(left), and its scatter plot, Fig. 15(right).

Appendix B

In the following we give an illustration of critical points for an arbitrary two-dimensional function defined on spherical surface, \(i.e., x \in \mathbb{S}^2\), and their respective location in scatter plot.
Fig. 1.
(left) Cortical thickness and surface coordinates. (right) For our processing, cortical thickness has been projected on to a sphere. Cortical thickness is measured in mm and color bar corresponds to thickness values at each vertex.
Fig. 2.
Weighted Fourier Series representation of cortical thickness signal. (top) Degree $k = 5, 20, 50$ and $70$ (left to right). (bottom) Bandwidth $\sigma = 0.01, 0.001, 0.0005$ and $0.00001$ (left to right).
Fig. 3.
Critical points pairing scheme for a one-dimensional function (based on [40]), also see Appendix A.
Fig. 4.
Flow chart to represent the complete algorithm.
Fig. 5.
(a) Scatter plot of an AD subject. The units on the Birth-Death axes are mms. In general, these scatter plots may have arbitrary units, (b) Concentration map of the scatter plot using gaussian kernel of bandwidth of 0.2 on a square grid $[0, 5]^2$ with 50$^2$ pixels (colorbar represent the density of points), (c) PDF of a concentration map.
Fig. 6.
Group analysis of extracted concentration maps for left and right hemispheres. Above are shown the differences in means for each pixel, (AD - control for left and right hemispheres resp.) Below are shown the $p$-values (negative log-scale). The differences are highly significant, ($p < 10^{-16}$) and suggest both gross and subtle anatomical variations are present in the data.
Fig. 7.
ROC curves demonstrating performance of topology-based representations of ADNI data.
Fig. 8.
Pearson’s correlation between classification confidence of algorithm and cognitive biomarkers.
Fig. 9.
The flat maps to show topologically interesting brain regions. Angles $\theta$ and $\phi$ (zenith and azimuthal respectively) are associated with spherical atlas. (Top) Differences in means at each vertex, (AD vs. control for left and right hemispheres respectively). (Bottom) $p$-values in negative log-scale.
Fig. 10.
Example: an arbitrary one-dimensional signal.
Fig. 11. Sublevel sets $R(y)$ (i.e., $f(x)$ in solid red below blue line) are drawn on ordered set $C$ for all critical points of $f(x)$, $C = \{ c_1 \leq c_2 \leq c_3 \leq c_4 \}$. One homological class (1-dimensional hole) is born at $c_1$, let us call this class $H_1$. Second class is born at $c_2$ – call this class $H_2$. $H_2$ class died at $c_3$ (i.e., the 1-dimensional hole opens up at $c_2$, closes off when entering $c_3$). $H_1$ died at $c_4$ (hole opened at $c_1$ and closed off at $c_4$).
Fig. 12.
Critical points pairing scheme to preserve the history of homological classes (*persistence*) present in $f(x)$. (a) Pair $(c_1; c_4)$ corresponds to the “life duration” of hole $H_1$. (b) Similarly, $(c_2; c_3)$ corresponds to the “life duration” of hole $H_2$. 
Fig. 13.
$f(x)$ and its persistence scatter plot.
Fig. 14.
Example: an arbitrary two-dimensional signal defined on $x \in \mathbb{R}^2$ and multiple views of its Weighted Fourier Series representation.
Fig. 15. (left) Flat map of WFS $f(x)$. (right) Persistence scatter plot. Arrow shows the spatial location of extremum points (birth-death) that correspond to single entry in scatter plot.
TABLE I
Demographic details and baseline cognitive status measures of the study population.

<table>
<thead>
<tr>
<th></th>
<th>AD (mean)</th>
<th>AD (s.d.)</th>
<th>Controls (mean)</th>
<th>Controls (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subjects</td>
<td>160</td>
<td>–</td>
<td>196</td>
<td>–</td>
</tr>
<tr>
<td>Age</td>
<td>75.53</td>
<td>7.41</td>
<td>76.09</td>
<td>5.13</td>
</tr>
<tr>
<td>Gender(M/F)</td>
<td>86/74</td>
<td>–</td>
<td>101/95</td>
<td>–</td>
</tr>
<tr>
<td>APoE4 carriers</td>
<td>76</td>
<td>–</td>
<td>81</td>
<td>–</td>
</tr>
<tr>
<td>MMSE at Baseline</td>
<td>21.83</td>
<td>5.98</td>
<td>28.87</td>
<td>3.09</td>
</tr>
<tr>
<td>Years of Education</td>
<td>13.81</td>
<td>4.61</td>
<td>15.87</td>
<td>3.23</td>
</tr>
</tbody>
</table>
TABLE II

A table of $p$-values on predicted neuropsychological scores.

<table>
<thead>
<tr>
<th>Neuropsychological Scores</th>
<th>p-value</th>
<th>Neuropsychological Scores</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rey Auditory Trial 1</td>
<td>$10^{-10}$</td>
<td>Trail Making A</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>Rey Auditory Trial 2</td>
<td>$\approx 0$</td>
<td>Trail Making B</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>Rey Auditory Trial 3</td>
<td>$10^{-16}$</td>
<td>Trail Making A &amp; B errors committed</td>
<td>$10^{-14}$</td>
</tr>
<tr>
<td>Rey Auditory Trial 4</td>
<td>$\approx 0$</td>
<td>Digit Span</td>
<td>$\approx 0$</td>
</tr>
<tr>
<td>Rey Auditory Trial 5</td>
<td>$\approx 0$</td>
<td>Boston Naming (Spontaneous)</td>
<td>$10^{-15}$</td>
</tr>
<tr>
<td>Rey Auditory Trial 6</td>
<td>$\approx 0$</td>
<td>Boston Naming (Cues Given)</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>Rey Auditory Trial B</td>
<td>$10^{-12}$</td>
<td>Boston Naming (Phonemic Cues Given)</td>
<td>$10^{-9}$</td>
</tr>
<tr>
<td>Category Fluency: Animals</td>
<td>$10^{-16}$</td>
<td>Boston Naming (Correct Response)</td>
<td>$10^{-8}$</td>
</tr>
<tr>
<td>Category Fluency: Vegetables</td>
<td>$\approx 0$</td>
<td>Boston Naming (Total Number Correct)</td>
<td>$10^{-15}$</td>
</tr>
<tr>
<td>Category Fluency: Vegetables, Intrusions</td>
<td>0.0002</td>
<td>ANART errors</td>
<td>$10^{-11}$</td>
</tr>
</tbody>
</table>
**TABLE III**

Measures of our method’s accuracy on classification experiments.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Hemisphere</td>
<td>73.31%</td>
<td>0.7955</td>
</tr>
<tr>
<td>Right Hemisphere</td>
<td>73.03%</td>
<td>0.7973</td>
</tr>
<tr>
<td>Averaged Confidence</td>
<td>75.00%</td>
<td>0.8063</td>
</tr>
</tbody>
</table>
TABLE IV

Predictive accuracies of unified kernels (ROI+Topology).

<table>
<thead>
<tr>
<th>Hemisphere</th>
<th>ROI Kernel</th>
<th>Accuracy</th>
<th>Kernel Combination</th>
<th>Accuracy</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>Hippocampus</td>
<td>79.03%</td>
<td>Hippocampus + Topology</td>
<td>83.09%</td>
<td>0.016</td>
</tr>
<tr>
<td>Left</td>
<td>Middle Temporal</td>
<td>78.13%</td>
<td>Middle Temporal + Topology</td>
<td>79.59%</td>
<td>0.336</td>
</tr>
<tr>
<td>Left</td>
<td>Inferior Temporal</td>
<td>79.88%</td>
<td>Inferior Temporal + Topology</td>
<td>82.51%</td>
<td>0.068</td>
</tr>
<tr>
<td>Right</td>
<td>Hippocampus</td>
<td>75.88%</td>
<td>Hippocampus + Topology</td>
<td>79.30%</td>
<td>0.045</td>
</tr>
<tr>
<td>Right</td>
<td>Middle Temporal</td>
<td>74.64%</td>
<td>Middle Temporal + Topology</td>
<td>78.13%</td>
<td>0.013</td>
</tr>
<tr>
<td>Right</td>
<td>Inferior Temporal</td>
<td>72.59%</td>
<td>Inferior Temporal + Topology</td>
<td>75.22%</td>
<td>0.054</td>
</tr>
</tbody>
</table>