CLASSIFICATION OF ALZHEIMER′S DISEASE USING UNSUPERVISED DIFFUSION COMPONENT ANALYSIS

Dominique Duncan

Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute
Keck School of Medicine of USC, University of Southern California, USA

Thomas Strohmer

Department of Mathematics
University of California, Davis, USA

and for the Alzheimer′s Disease Neuroimaging Initiative∗

Abstract. The goal of this study is automated discrimination between early stage Alzheimer′s disease (AD) magnetic resonance imaging (MRI) and healthy MRI data. Unsupervised Diffusion Component Analysis, a novel approach based on the diffusion mapping framework, reduces data dimensionality and provides pattern recognition that can be used to distinguish AD brains from healthy brains. The new algorithm constructs coordinates as an extension of diffusion maps and generates efficient geometric representations of the complex structure of the MRI data. The key difference between our method and others used to classify and detect AD early in its course is our nonlinear and local network approach, which overcomes calibration differences among different scanners and centers collecting MRI data and solves the problem of individual variation in brain size and shape. In addition, our algorithm is completely automatic and unsupervised, which could potentially be a useful and practical tool for doctors to help identify AD patients.

1. **Background.** Alzheimer′s disease (AD), the most common type of dementia, currently affects approximately 5.2 million people in the US, with a significant increase predicted in the near future. Over 35 million people worldwide are living with AD; this number is expected to double by 2030 and more than triple by 2050 to 115 million [1]. In AD patients, neurons along with their connections are progressively destroyed, leading to loss of cognitive function and eventually death [15]. Therapeutic intervention is generally considered more likely to be beneficial in the early stages of the disease. Thus, it is extremely important to identify the disease as early as possible in order to administer treatments that will effectively stop the disease.

Mild Cognitive Impairment (MCI), a transitional stage between normal aging and the development of dementia, has been defined to account for the intermediate cognitive state where patients are impaired on one or more standardized cognitive tests but do not meet the criteria for clinical diagnosis of dementia [10]. MCI has attracted increasing attention lately since it offers an opportunity to target the disease process early.

Neuroimaging has been shown to be a powerful tool for studying changes in the progression of AD as well as therapeutic efficacy in AD patients. Magnetic resonance imaging (MRI) scans can reveal features that are predictive of a patient developing AD. Our goal is to use these features to distinguish brains of patients in early stages of AD from brains of healthy patients.

A novel approach based on the diffusion map framework is used [3]; diffusion mapping provides dimensionality reduction of the data as well as pattern recognition that can be used to distinguish AD brains from non-AD brains. A new algorithm, Unsupervised Diffusion Component Analysis, which is an extension of diffusion maps, constructs coordinates that generate efficient geometric representations of the complex structures in the MRI. The diffusion map approach has been effective in other classifications using brain data, in particular, preseizure states of patients with epilepsy [4]. Diffusion maps have also been effective in classifications in various nonmedical areas, such as finance and military applications.

There have been other studies on classifying AD and non-AD patients; some of them use principal components analysis (PCA) or independent component analysis (ICA). Recently more work has been done using multivariate approaches rather than the traditional voxel-by-voxel approach [5]. However, the key difference between our method and other methods that have been used to classify and detect onset of AD in early stages is the nonlinear and local network approach, which is necessary for eliminating the calibration differences of MRI of patients with different shapes and sizes of brains as well as different scanners and centers collecting data. Furthermore, another major difference and improvement in our algorithm is that it is completely automatic and unsupervised, which could potentially be an incredibly useful tool for doctors to help identify AD patients.

2. **Data.** Data used in the preparation of this article were obtained from the Alzheimer′s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations, as a 60 million, 5-year public-private partnership. The Principle Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. Presently, more than 800 participants, aged 55 to 90 years, have been recruited from over 50 sites across the United States and Canada, including approximately 200 cognitively normal older individuals (i.e., healthy controls or HCs) to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. Baseline and longitudinal imaging, including structural MRI scans collected on the full sample and PIB and FDG PET imaging on a subset are collected every 612 months. Additional baseline and longitudinal data including other biological measures (i.e. cerebrospinal fluid (CSF) markers, APOE
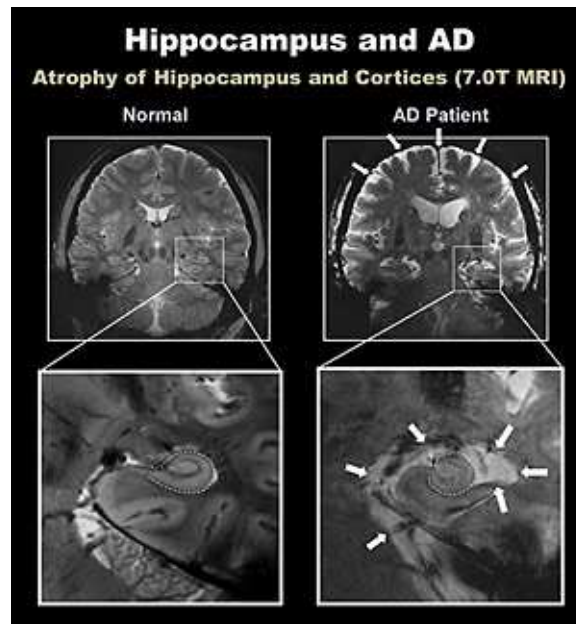
FIGURE 1. An example of a normal MRI and an AD MRI, showing differences in the hippocampal region

and full-genome genotyping via blood sample) and clinical assessments including neuropsychological testing and clinical examinations are also collected as part of this study. Written informed consent was obtained from all participants and the study was conducted with prior institutional review board's approval. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer′s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. For further and updated information, see www.adni-info.org.

3. **Methods.** We assume that the features differentiating patients with AD are represented in the MRI data. We would like to detect these features and distinguish brains of patients in the early stages of AD from brains of non-AD patients.

Figure 1 shows an example of a normal MRI and an AD MRI; sometimes it is not straightforward to identify such small changes in the images, so it would be useful to have an automatic way to identify AD patients using only structural MRI. Figure 2 is another example that shows the MRI of 3 different 75 year old patients: normal, MCI, and AD.

Diffusion maps [3] have been a useful tool in reducing the dimensionality of the data as well as providing a measure for pattern recognition and feature detection. Since diffusion mapping may detect special features in the data, it can be used to determine differences in brains of patients with AD compared to normal, healthy

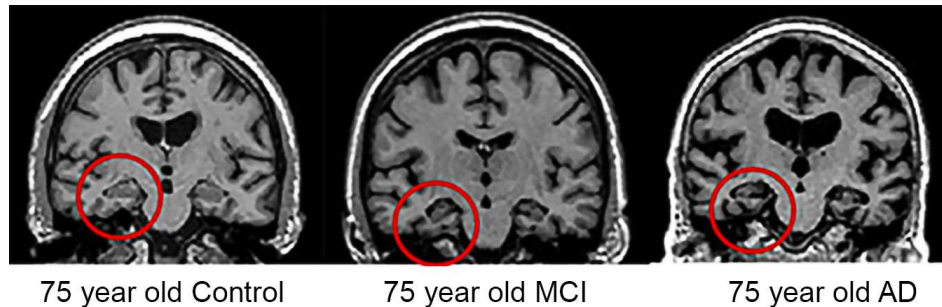75 year old Control   75 year old MCI   75 year old AD

FIGURE 2. An example of 3 different MRI: 75 year old control, 75 year old MCI, and 75 year old AD

brains. However, diffusion maps assume access to the process that they aim to classify. In MRI data, the relationship between the pixels of the images and the underlying brain activity may be stochastic, and the data are assumed to be noisy due to the calibration. Hence, diffusion mapping is not the most effective direct approach to use with MRI data. A recently developed algorithm, which is an extension of diffusion maps, may be more applicable in the case of classifying AD [12, 2]. This new algorithm assumes a stochastic mapping between the underlying processes and the measurements, so the mapping is inverted, and a kernel is used to recover the underlying activity [12]. Thus it seems that this proposed algorithm is more appropriate than diffusion maps for our data.

We introduce an algorithm that relies on the work by Talmon and Coifman [12] to extract the underlying brain structure from the MRI. The algorithm is an extension of diffusion maps and uses local PCA [9]. PCA is another dimensionality reduction method, in which the goal is to compute the most meaningful basis to re-express a large and noisy dataset. This new basis can reveal hidden patterns and structure in the data as well as remove the noise. An orthogonal linear transformation converts the data to a new coordinate system for more effective analysis. The largest variance in the data is represented by the first coordinate or the first principal component. An important difference between the proposed algorithm and PCA is the use of nonlinear local analysis in the extension as opposed to PCA, which assumes the linear global information of the data. For the MRI data, we perform PCA on local regions of the images and then integrate the local information using a kernel and obtain a single model for all of the data. We use a data-driven adapted distance between blocks of MRI to approximate the Euclidean distance between the features from the MRI that are considered noisy due to calibration differences.

The MRI data form 3D matrices, because the scanner records 2D slices of the brain. Slices cannot be considered in isolation because of variance in their number and thickness across different scanners and scanning protocols. The full brain 3D matrices are subdivided into vectors that are composed of overlapping neighborhoods around pixels of size 8x8x8, and these submatrices are overlapped by 50% for smoothing purposes and to account for the fact that our submatrix size may split a particular brain structure that we would prefer remain whole. This overlapping is natural from the nonlinear assumptions in the approach. These submatrices are reshaped into vectors of length 512 (8x8x8). Then the vectors from the MRI of patients with AD are compared to the vectors from the MRI of healthy patients to determine if certain features are different and can be used to identify AD.

For each set of feature vectors for the 4 MRI datasets that we consider, we compute histograms using 20 bins to approximate the probability distributions, because the MRI data are assumed to be stochastic from various effects. After combining the results for the 4 MRI, we calculate the Earth Mover's Distance [11] rather than computing Euclidean distances between pixels or between boxes. This is a method to evaluate dissimilarity between multi-dimensional distributions in some feature space where a distance measure between single features is given. The Earth Mover's Distance is called the Wasserstein metric in optimal transport where the problem is to transport a mass from one location to another. Using this method in our algorithm is useful, because it naturally extends the notion of a distance between elements to that of a distance between sets of elements. Furthermore, it is applicable to MRI data, because it allows for partial matches in a natural way, which helps to deal with occlusions and clutter in image retrieval applications.

To reduce the chance of bias in the construction, we introduce a random shuffle in the columns of the matrix composed of feature vectors and apply a random projection as a method to reduce the large amount of data. Then we apply the Discrete Cosine Transform [13]. If the data are uncorrelated, we expect to obtain some approximation of a delta function with a spike at the origin after applying the Discrete Cosine Transform.

Given one of these feature vectors, $\mathbf{S}_y(m)$, we compute the empirical local covariance matrix Σ_m within a fixed interval, J,

$$\Sigma_m = \frac{1}{J} \sum_{m'=m-J+1}^{m} (\mathbf{S}_y(m') - \mu_m)(\mathbf{S}_y(m') - \mu_m)^T, \tag{1}$$

where μ_m is the empirical local mean of the feature vectors in the interval, and m describes the data that have been classified in cells by a histogram.

The dynamics of the controlling factors from the data are described by normalized independent Ito processes described in the stochastic differential equation below:

$$d\theta_i(t) = a_i(\theta_i(t))dt + dw_i(t), \tag{2}$$

where $i = 1, 2, ..., d$. $(a_1, ..., a_d)$ in the above equation are (possibly nonlinear) unknown drift coefficients and $w = (w_1, ..., w_d)$ is a $d-$dimensional independent white noise. An n-dimensional process $(Y(t), t \geq 0)$ is the observation and a noisy measurement process Z arises as $Z(t) = g(Y(t), V(t))$, where V is a stationary noise process with unknown distribution.

We define a nonsymmetric distance known as the Mahalanobis distance using the covariance matrices, a_Σ^2, and a symmetric distance d_Σ^2. Mahalanobis distances between empirical distribution estimators (e.g., histogram vectors) are used to construct the affinity measure between segments in the series. Then anisotropic kernels are constructed and diffusion maps are applied to obtain a low-dimensional embedding, which uncovers the intrinsic representation. It has been shown in [3] that this distance approximates the Euclidean distance between the underlying factors in the data by local linearization of the nonlinear transformation. These distances, between points m and m' in the dataset M, are defined as follows:

$$a_\Sigma^2(m, m') = (\mathbf{S}_y(m) - \mathbf{S}_y(m'))^T \Sigma_{m'}^{-1} (\mathbf{S}_y(m) - \mathbf{S}_y(m')), \tag{3}$$

$$d_\Sigma^2(m, m') = \frac{1}{2}(a_\Sigma^2(\mathbf{S}_y(m), \mathbf{S}_y(m')) + a_\Sigma^2(\mathbf{S}_y(m'), \mathbf{S}_y(m))). \tag{4}$$

We are able to recover these underlying factors using an eigendecomposition of an appropriate Laplace operator (kernel). A kernel is used to compare the underlying

factors, and ϵ is the kernel scale set according to the Mahalanobis distance. This kernel is used to define the local geometries of the graph between m and m' from the dataset M.

We construct an NxN nonsymmetric affinity matrix A, whose (m, m') element is given by

$$A^{m,m'} = exp-\frac{a_{\sum}^2(\mathbf{S}_y(m), \mathbf{S}_y(m'))}{\epsilon} \tag{5}$$

where $\epsilon > 0$ is the kernel scale that is calculated by taking the median of all pairwise distances of the original data matrix.

The matrix formed from the elements with the above exponential converges to a low dimensional manifold and the eigenvectors parametrize the underlying structures in the data.

The kernel is normalized by a diagonal density matrix, which enables us to consider the sampling as uniform. The normalized matrix can be viewed as a Markov transition probability matrix for a jump process over the measurements. We then define an NxN symmetric matrix W as

$$W^{m,m'} = \sum_{r \in \mathcal{R}} A^{m,r} A^{m',r}. \tag{6}$$

Then an eigendecomposition is performed to address the nonuniform sampling of the data. The ℓ eigenvectors found from the eigendecomposition corresponding to the few largest eigenvalues provide a parametrization of the features, allowing for significant data dimensionality reduction and capturing the features that may identify patients with AD.

$$\mathbf{S}_y(m) \mapsto [\psi_1(m), \psi_2(m), ..., \psi_\ell(m)]^T, \tag{7}$$

where $\psi_i(m)$ is the ith eigenvector. To determine which eigenvectors to use for this classification problem, we pick the optimal eigenvector embedding with a computable, reproducible criterion instead of visual inspection. All possible combinations of 3 or 4 eigenvectors are considered. We compute the center of mass of the new embedded points. Then to choose which embedding provides the best separation with AD points separated from the rest of the embedded points, we calculate the variance of all points in the embedding that correspond to the normal MRI data to that center of mass. The variance of the normal points is divided by the variance of all points in the embedding that correspond to the AD MRI data to the center of mass for each case. We choose the maximum variance ratio and consider the top 3 cases and choose those sets of eigenvectors.

The details are summarized in the following table with algorithmic listing.

Algorithm

1: Obtain MRI data of n brains,
2: Partition each 3-dimensional matrix of data into overlapping submatrices,
3: Reshape each small submatrix into a vector; place each vector side by side to form a matrix,
4: Compute histograms (along matrix columns) using 20 bins,
5: Calculate the Earth Mover's Distance between consecutive feature vectors,
6: To reduce the chance of bias, introduce a random shuffle in the columns of the matrix and apply a random projection,
7: Apply the Discrete Cosine Transform,
8: Calculate local covariance matrices for overlapping windows,
9: Compute the eigenvalue decomposition to obtain eigenvalues and corresponding eigenvectors,
10: Calculate inverse covariance matrices to calculate the Mahalanobis Distance,
11: Use the median of all pairwise distances of the data matrix to choose epsilon, the Gaussian kernel scale,
12: Compute the affinity matrix and build a Gaussian kernel according to (5),
13: Normalize the kernel by a diagonal density matrix and employ eigenvalue decomposition to obtain the eigenvalues and eigenvectors,
14: Consider all possible combinations of 3 or 4 eigenvectors for the embeddings; compute the center of mass for each embedding as well as the variance of the embedded points (specifically, the ratio of the variance of the normal points divided by the variance of the AD points) to determine the optimal embedding.

---

4. **Results.** Initially, using the algorithm to compare 2 AD and 2 normal brains, we found a distinct separation, as shown in Figure 3. We decided to analyze 10 examples, in which there is one different AD MRI in each example and the same three normal MRI. This discrimination would be beneficial for doctors to identify AD patients, because they could use a reference dataset of normal MRI data and compare individual patient MRI data against this dataset. For each of these 10 cases, we produced the embeddings of all combinations of 3 eigenvectors, for example, Figure 5. One example of this is Figure 4. In that figure, the large green dot represents the center of mass of all of the points in the embedding, and this is used to calculate the variance of the other points in the embedding.

From all iterations of possible combinations of 3 eigenvectors, we select the top 5 embeddings that produce the best separation for the AD points and show that each time, our automatic and unsupervised algorithm is able to select as the best embedding one of these top 5 options by checking the variance ratio (variance of normal points divided by variance of AD points from the center of mass in the embedding), displayed in Figure 6. We also checked all combinations of 4 eigenvectors and plotted the variance ratio, as in Figure 7 with similar results. Furthermore, we were able to trace back the embeddings to the original data to determine which areas in the brain seem to be most differentiating between healthy and AD data, and we found these areas to be located in the temporal lobe.

5. **Discussion.** A method similar to the one proposed in this paper has already proved to be effective in identifying preseizure states in intracranial EEG data by
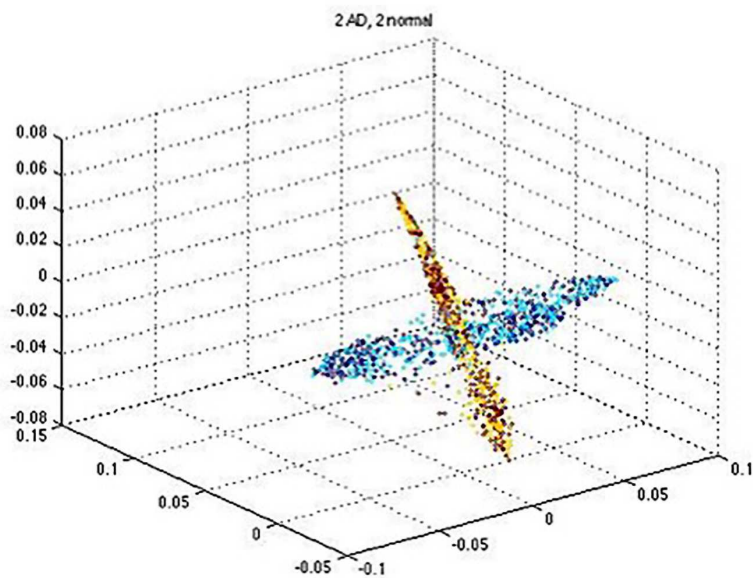
FIGURE 3. An example using 4 different MRI (2 normal and 2 AD) of one embedding using 3 eigenvectors, and each color represents a different MRI: blue are normal; yellow and red are AD.
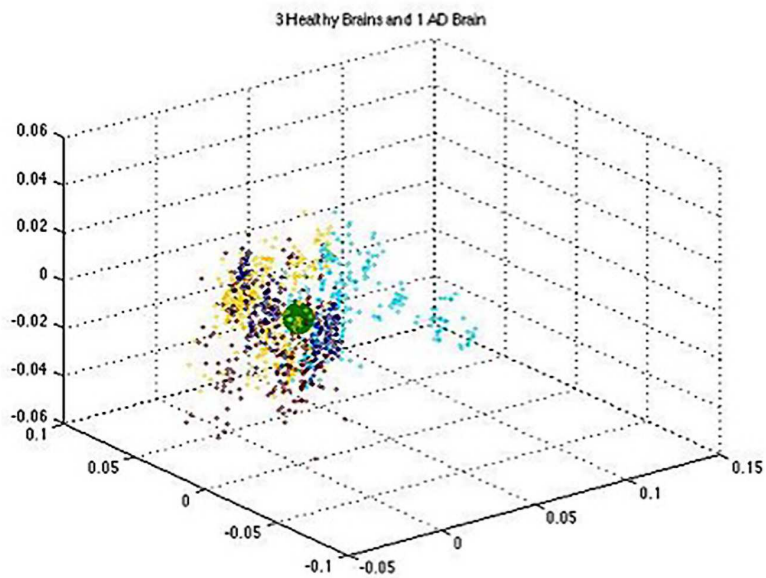

FIGURE 4. An example using 4 different MRI (3 normal and 1 AD) of one embedding using 3 eigenvectors, and each color represents a different MRI: light blue is AD.
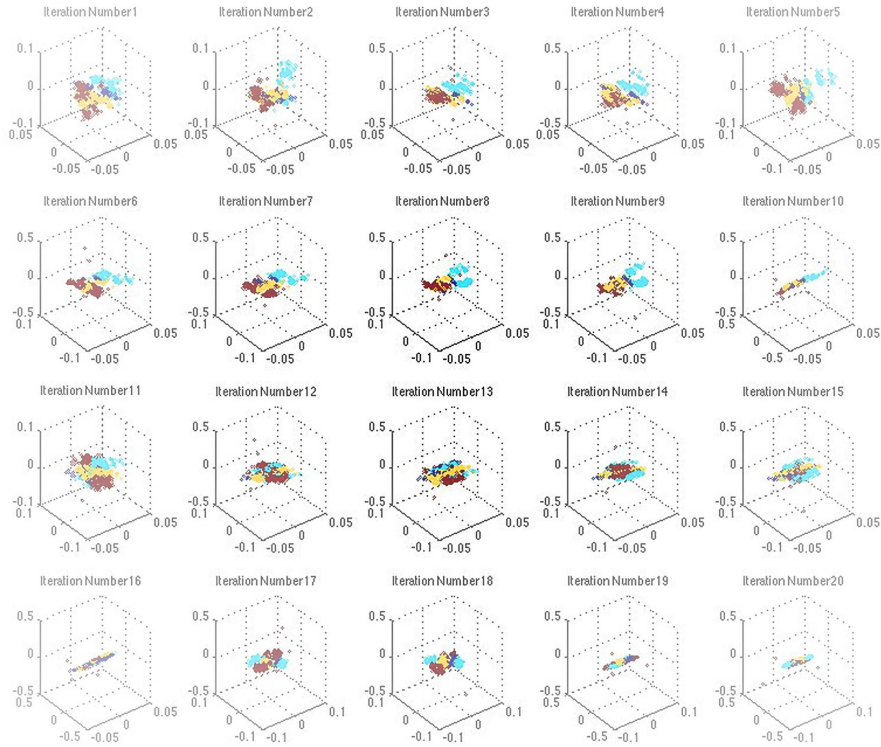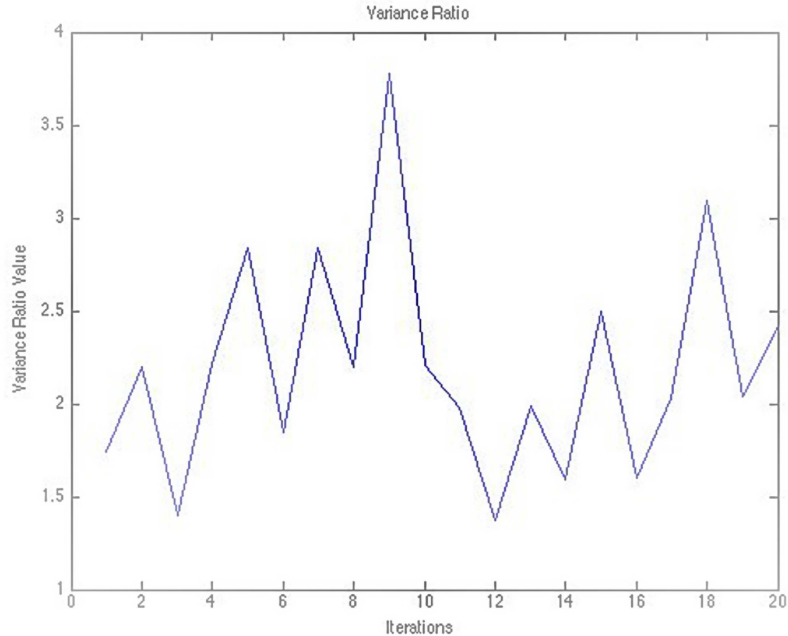
FIGURE 5. An example using 4 different MRI (3 normal and 1 AD) of all embeddings with various combinations of 3 eigenvectors representing the axes, and each color represents a different MRI: light blue is AD.

providing a distinction between interictal (period between seizures) and preseizure states of a patient with epilepsy [4].

Other studies that have focused on identifying and classifying AD patients have used multivariate techniques, because they have attractive features that cannot be discovered by the more commonly used univariate, voxel-wise, techniques [5].

ICA based methods have been used for analyzing neuroimaging data, such as MRI data. Yang et al. [14] used ICA and a support vector machine (SVM) to classify AD MRI data. They first aligned and normalized all MRI scans studied using statistical parametric mapping. Next, ICA was applied to the images to extract features used for classification. The SVM was then used to classify the images based on the independent component coefficients.

6. **Conclusions.** Unsupervised Diffusion Component Analysis, a novel algorithm which combines diffusion maps and PCA with other techniques, is used to study the differences between healthy and AD patients. The extensions lead to efficiency
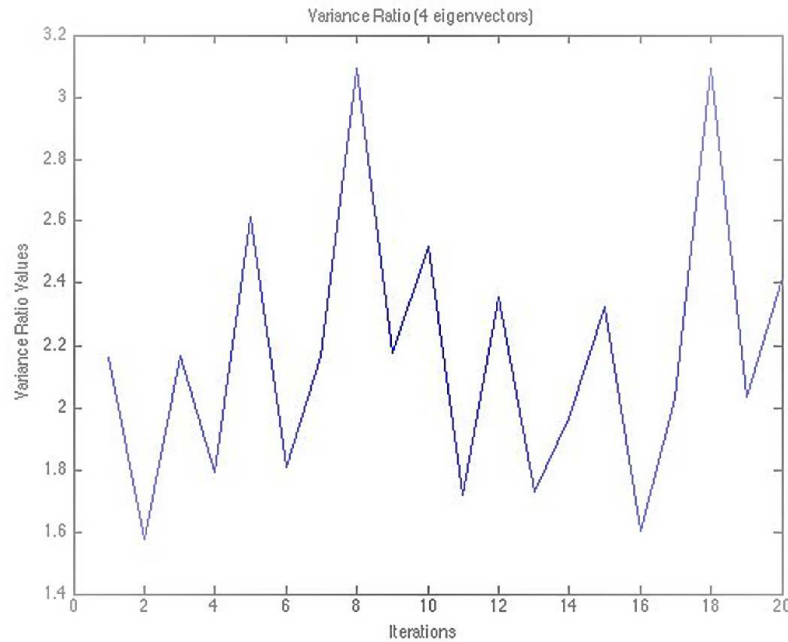
FIGURE 6. An example using 4 different MRI (3 normal and 1 AD) of the Variance Ratio for all embeddings with various combinations of 3 eigenvectors

in use, in terms of reduced computational complexity, which have the potential to become useful techniques for practitioners in the field.

The key difference between our method and others used to classify and detect AD early in its course is our nonlinear and local network approach, which overcomes calibration differences among different scanners and centers collecting MRI data and solves the problem of individual variation in brain size and shape. Additionally, our algorithm is completely automatic and unsupervised, which could potentially be a very useful tool for doctors to help identify AD patients. Furthermore, we have tried to address some disadvantages with multivariate approaches, such as the higher demands of computational and mathematical literacy on the data analyst. After the initial work of developing this algorithm and determining a reference bank of healthy/normal brains, the remaining analysis is kept straightforward so that Unsupervised Diffusion Component Analysis could present a simple tool for doctors to use in diagnosing Alzheimer′s Disease.

Future work will include testing on a larger sample size as well as testing on data from patients with mild cognitive impairment to see if the algorithm is able to separate that data from the data of healthy patients, which would allow doctors to diagnose patients prior to AD onset.

FIGURE 7. An example using 4 different MRI (3 normal and 1 AD) of the Variance Ratio for all embeddings with various combinations of 4 eigenvectors

REFERENCES

[1] Alzheimer′s Association: Alzheimer′s disease facts and figures. *Alzheimer′s & Dementia,* **9** (2013), 208–245.

[2] N. Ahmed, T. Natarajan and K. R. Rao, Discrete cosine transform, *IEEE Transactions on Computers,* **23** (1974), 90–93.

[3] R. R. Coifman and S. Lafon, Diffusion maps, *Appl. Comp. Harm. Anal.,* **21** (2006), 5–30.

[4] D. Duncan, R. Talmon, H. P. Zaveri and R. R. Coifman, Identifying preseizure state in intracranial EEG data using diffusion kernels, *Math Biosci Eng,* **10** (2013), 579–590.

[5] C. Habeck and Y. Stern, Alzheimer′s disease neuroimaging initiative, Multivariate data analysis for neuroimaging data: Overview and application to Alzheimer's disease, *Cell Biochem Biophys.,* **58** (2010), 53–67.

[6] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli and J.-P. Thiran, Mapping human whole-brain structural networks with diffusion MRI, *PLoS ONE,* **2** (2007), e597.

[7] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen and O. Sporns, Mapping the structural core of human cerebral cortex, *PLoS Biol,* **6** (2008), e159.

[8] S. Norton, F. E. Matthews, D. Barnes, K. Yaffe and C. Brayne, Potential for primary prevention of Alzheimer's disease: an analysis of population-based data, *Lancet Neurology,* **13** (2014), 788–794.

[9] C. Syms, Principal components analysis, *Reference Module in Earth Systems and Environmental Sciences Encyclopedia of Ecology,* (2008), 2940–2949.

[10] R. C. Petersen, Mild cognitive impairment clinical trials, *Nature Reviews Drug Discovery*, **2** (2003), 646–653.

[11] Y. Rubner, C. Tomasi and L. J. Guibas, A metric for distributions with applications to image databases, *IEEE 6th International Conference on Computer Vision,* (1998), 59–66.

[12] R. Talmon and R. R. Coifman, Differential stochastic sensing: intrinsic modeling of random time series with applications to nonlinear tracking, *PNAS,* (2012), 1–14.

[13] R. Talmon, D. Kushnir, R. R. Coifman, I. Cohen and S. Gannot, Parametrization of linear systems using diffusion kernels, *IEEE Transactions on Signal Processing,* **60** (2012), 1159–1173.

[14] W. Yang, R. L. Lui, J. H. Gao, T. F. Chan, S. T. Yau, R. A. Sperling and X. Huang, Independent component analysis-based classification of Alzheimer′s disease MRI data, *J. Alzheimers Dis,* **24** (2011), 775–783.

[15] J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo and V. A. Narayan, Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data, *BMC Neurology,* **12** (2012), 1–12.

E-mail address: dduncan@loni.usc.edu
E-mail address: strohmer@math.ucdavis.edu