Identifying Changes in Brain MRI in Early Stages of Alzheimer's Disease

Dominique Duncan

UC Davis

Micro and Macro Systems in Life Sciences

June 12, 2015

Outline

- Alzheimer's Disease background
- MRI Data
- Analysis Methods
- Results on classification
- Discussion and Future Work

Alzheimer's Disease

- Most common type of dementia
- Neurons along with their connections are progressively destroyed, leading to loss of cognitive function and eventually death
- Therapeutic intervention is most likely beneficial in early stages
- Mild Cognitive Impairment (MCI): transitional stage between normal aging and development of dementia
- Neuroimaging: powerful tool for studying changes in Alzheimer's Disease (AD) progression and therapeutic efficacy in AD patients
- Magnetic Resonance Imaging (MRI) scans are useful for identifying features that can help predict which patients will develop AD

Alzheimer's Facts

It's the only cause of death in the top 10 in America that CANNOT BE PREVENTED, CURED OR SLOWED.

ALMOST TWO THIRDS of Americans with Alzheimer's disease are women.

SENIORS dies with Alzheimer's or another dementia.

Only **45%** of people with **ALZHEIMER'S** disease or their caregivers report **BEING TOLD OF THEIR DIAGNOSIS.**

More than **90%** of people with the four most common types of **CANCER** have been **TOLD OF THEIR DIAGNOSIS.**

By 2050, these costs could rise as high as **$1.1 TRILLION.**

In 2015, Alzheimer's and other dementias will cost the nation **$226 BILLION.**

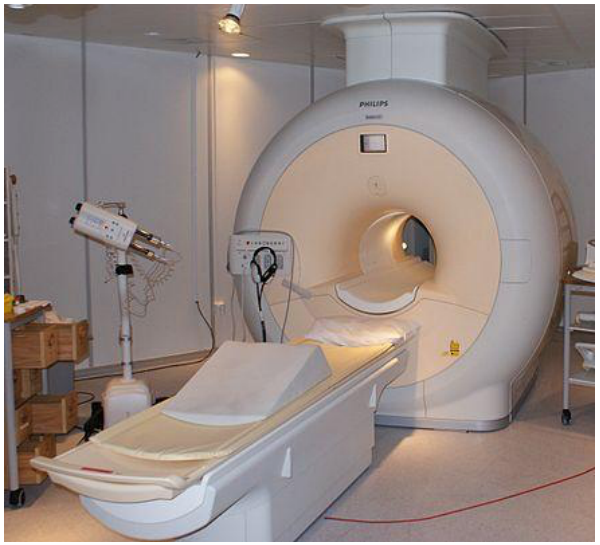
Alzheimer's Disease Facts

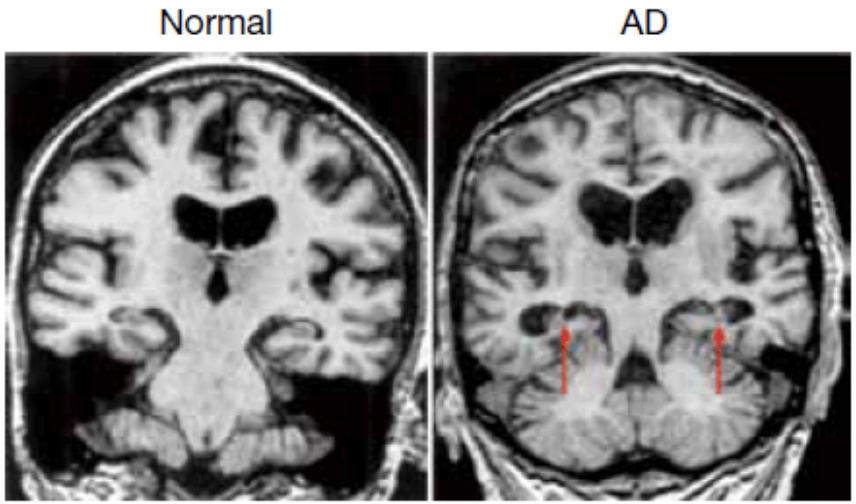
- Nearly 36 million people have AD or a related dementia worldwide (5.2 million Americans)
- More than 13 million Americans are expected to have AD by 2050
- AD and dementia is most common in Western Europe (then North America), least prevalent in Sub-Saharan Africa
- 1 in 4 people with AD have been diagnosed
- Early-onset AD (5%) can develop in people as young as age 30
- AD is a growing epidemic worldwide and the only leading cause of death that is still on the rise
- The cost of caring for AD patients in the US is estimated to be $220 billion per year ($605 billion worldwide, equivalent to 1% of the entire world's GDP)
- Over 15 million Americans are unpaid caregivers for someone with AD
- Typical life expectancy after an AD diagnosis if 4-8 years

(Alzheimer's Association Facts and Figures, 2014)

MRI Data

- Assumption: there are certain features in the brain images of patients with AD
- Main idea: we would like to discover these features to distinguish brains of patients in early stages of AD from brains of healthy patients
- Goal: to find those low dimensional features based on the observable high dimensional data
- Challenges: noisy data, unknown model

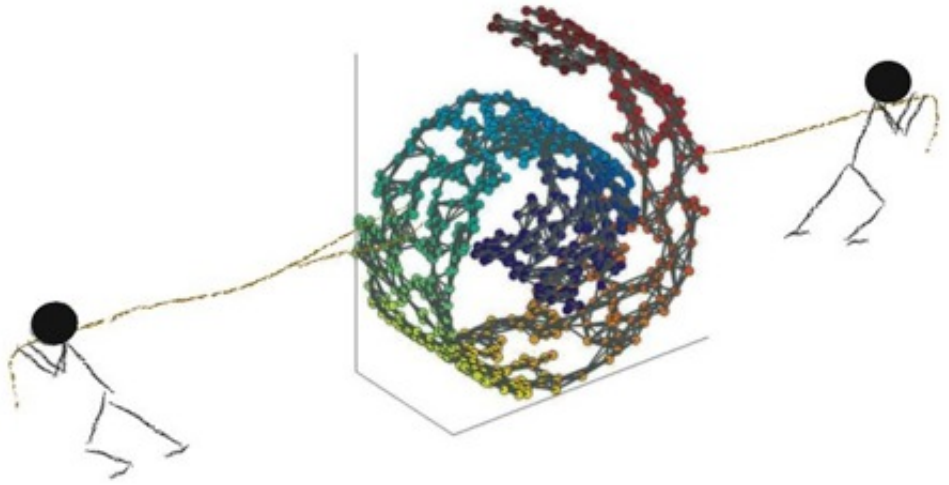




Compliment of Scanning Department, St. Teresa's Hospital

Manifold Learning

- Represent the data as points in a high dimensional space
- The points lie on a low dimensional structure (manifold) that is governed by latent factors
- Traditional techniques:
  - Laplacian eigenmaps [Belkin & Niyogi, 2003]
  - Diffusion maps [Coifman & Lafon, 2005; Singer & Coifman, 2008]



Diffusion Maps

- Recorded high dimensional data that measures brain activity and features
- We assume that for the problem of identifying AD, there are a few brain activity factors that distinguish AD brains from healthy brains
- Our goal is to find those controlling low dimensional factors based on the observable high dimensional data

Diffusion Maps

- Eigenfunctions of Markov matrices are used to construct coordinates that generate efficient representations of complex geometric structures
- Not only does diffusion mapping allow for dimensionality reduction of the data, but this method also provides pattern recognition so that specific parts of the data may be analyzed more closely
- Diffusion maps: nonlinear and local, unlike Principal Components Analysis (linear and global)

(Coifman, 2006)

General Outline

- For each measurement in the set, we compute the local histogram and covariance
- Construct an NxN symmetric affinity matrix (kernel) between the measurements
- Normalize the kernel to obtain a Laplace operator [Chung, 1997]
- The leading eigenvectors from the spectral decomposition represent the underlying factors

Intrinsic Modeling

- The mapping between the observable data and the underlying processes is often stochastic and contains measurement noise
  - Repeated observations of the same phenomenon usually yield different measurement realizations
  - The measurements may be performed using different instruments/sensors
- Each set of related measurements of the same phenomenon will have a different geometric structure
  - Depending on the instrument and the specific realization
  - Poses a problem for standard manifold learning methods

Methods

- There is noise in MRI data and issues with calibration; our method allows us to separate the signal from the noisy data.
- Instead of looking at the individual realizations (Diffusion Maps), we want to analyze the variability of the statistics
- We use local covariance matrices as feature vectors for the statistics and measure the variability using the Mahalanobis distance between covariances
- It can be shown that the combination of the two yields a local metric that is invariant to added noise

Mahalanobis Distance

- Measures the distance of a point x from a data distribution
- The data distribution is characterized by a mean and the covariance matrix, thus is hypothesized as a multivariate gaussian
- Used in pattern recognition as a similarity measure between the pattern (data distribution of training example of a class) and the test example)
- The covariance matrix gives the shape of how data are distributed in the feature space

Data Features and Metric

- The 3D matrices that are formed from the MRI are subdivided into vectors that are made up of overlapping neighborhoods around pixels
- These vectors from MRI of patients with AD are compared to the vectors from MRI of healthy patients
- We view the local histograms as feature vectors for each measurement, $s_y(m)$
- We compute histograms to approximate the probability distributions, because the MRI data are assumed stochastic from various effects

Data Features and Metric

- Given a feature vector $s_y(m)$, we compute the local covariance matrix in an interval of length J:

$$\boldsymbol{\Sigma}_m = \frac{1}{J} \sum_{m'=m-J+1}^{m} (\mathbf{s}_y(m') - \boldsymbol{\mu}_m)(\mathbf{s}_y(m') - \boldsymbol{\mu}_m)^T$$

- Mahalanobis distance:

$$a^2_\Sigma(m, m') = (\mathbf{s}_y(m) - \mathbf{s}_y(m'))^T \boldsymbol{\Sigma}_m^{-1} (\mathbf{s}_y(m) - \mathbf{s}_y(m'))$$

$$d^2_\Sigma(m, m') = (1/2)(a^2_\Sigma(m, m') + a^2_\Sigma(m', m)$$

-invariant under linear transformations, thus by lemma, noise resilient (Talmon & Coifman, 2013)

-approximates the Euclidean distance between samples of the underlying process

1) Training Stage

$$W_R \in R^{NxN}$$

$$W_{\mathcal{R}}^{m,m'} = \exp\left\{-\frac{d_{\Sigma}^2(m,m')}{\epsilon}\right\}$$

- ε is the kernel scale set according to the Mahalanobis distance
- If m and m' are in the same state, this is proportional to the probability that m and m' are in the same Gaussian state

(Kushnir, 2012; Talmon, 2012)

2) Test Stage

$$A \in R^{MxN}$$

$$A^{mm'} = \exp\left(-\frac{a_{\Sigma}^{2}(m,m')}{\epsilon}\right)$$

$$W_R = A^T A \rightarrow \{\lambda_i, \varphi_i\}$$

$$W = AA^T \rightarrow \{\lambda_i, \psi_i\}$$

After constructing a Gaussian kernel over our dataset, we normalize the kernel using the weighted graph Laplacian normalization to become a Markov transition matrix.
That matrix is conjugate to a symmetric matrix, and we calculate its spectral decomposition.
The eigenvalues of both matrices are identical.
We use the eigenvectors from the test stage for our data: any new datapoint, we embed into the previous graph

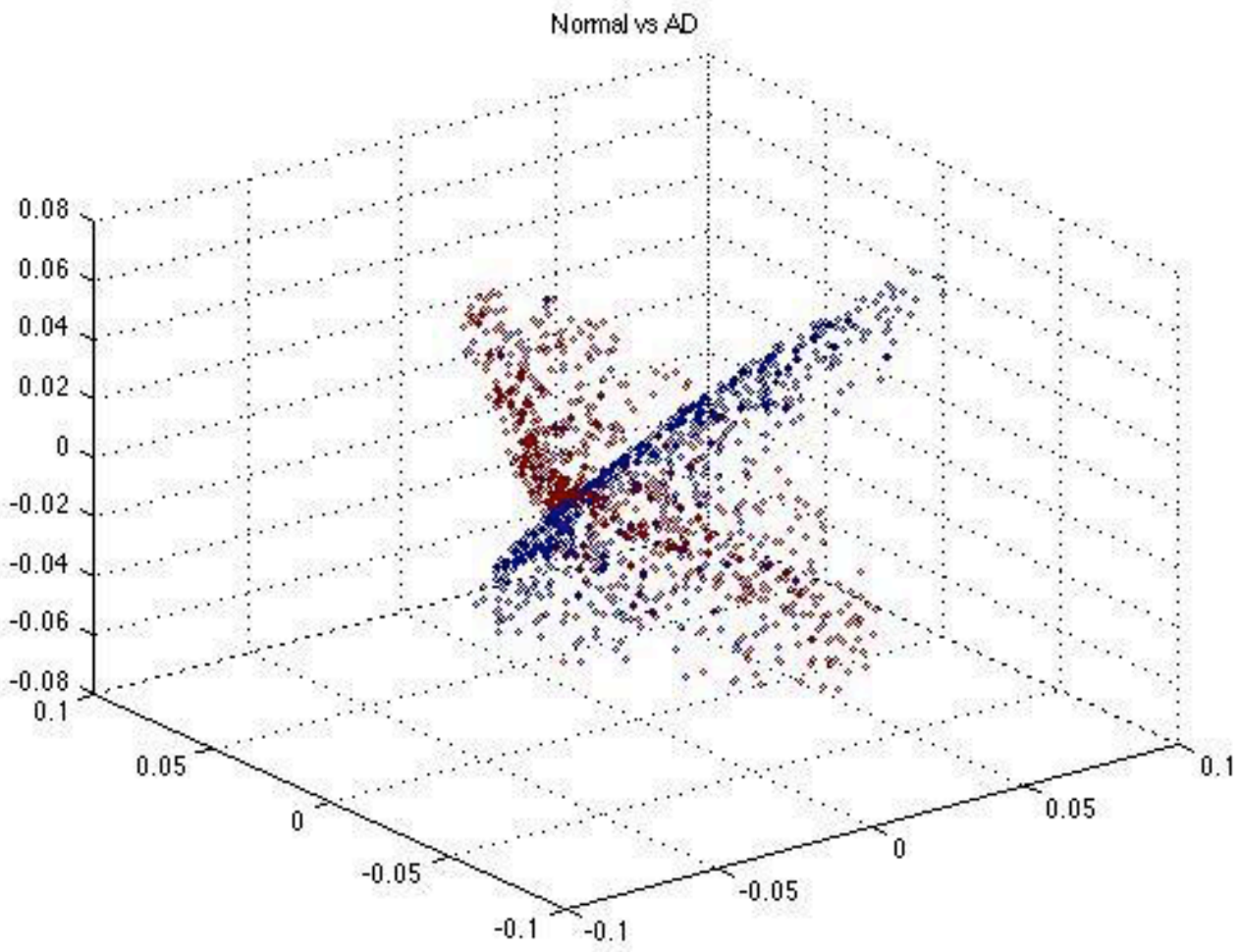
Eigendecomposition

$$\psi_i = \frac{1}{\sqrt{\lambda_i}} \varphi_i$$

$$\mathbf{s}_y(m) \mapsto [\lambda_1 \psi_1(m), \lambda_2 \psi_2(m), ..., \lambda_\ell \psi_\ell(m)]^T$$

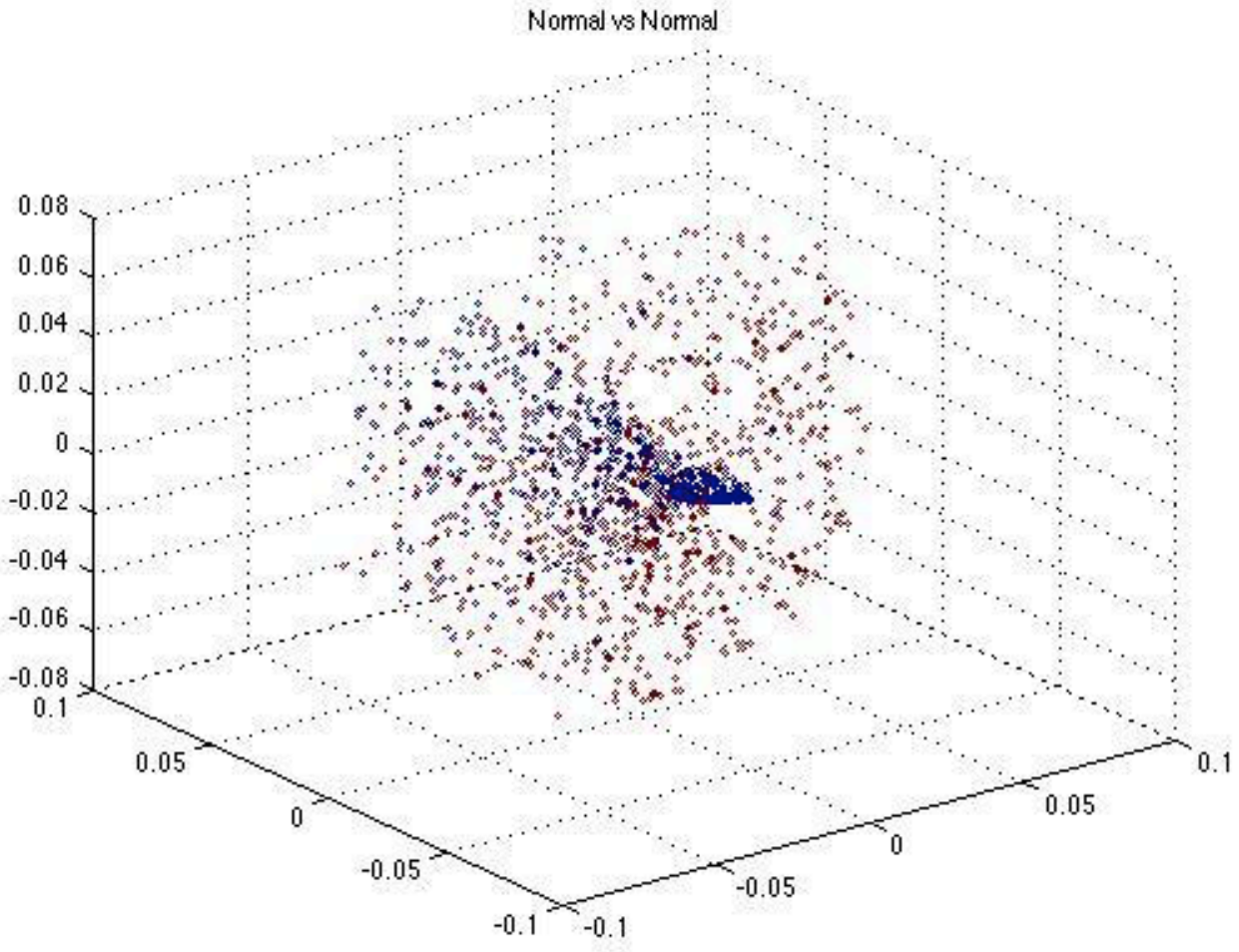
Analytical construction that allows us to embed the output of the measurements into the same normalized diffusion map embedding

AD vs Normal



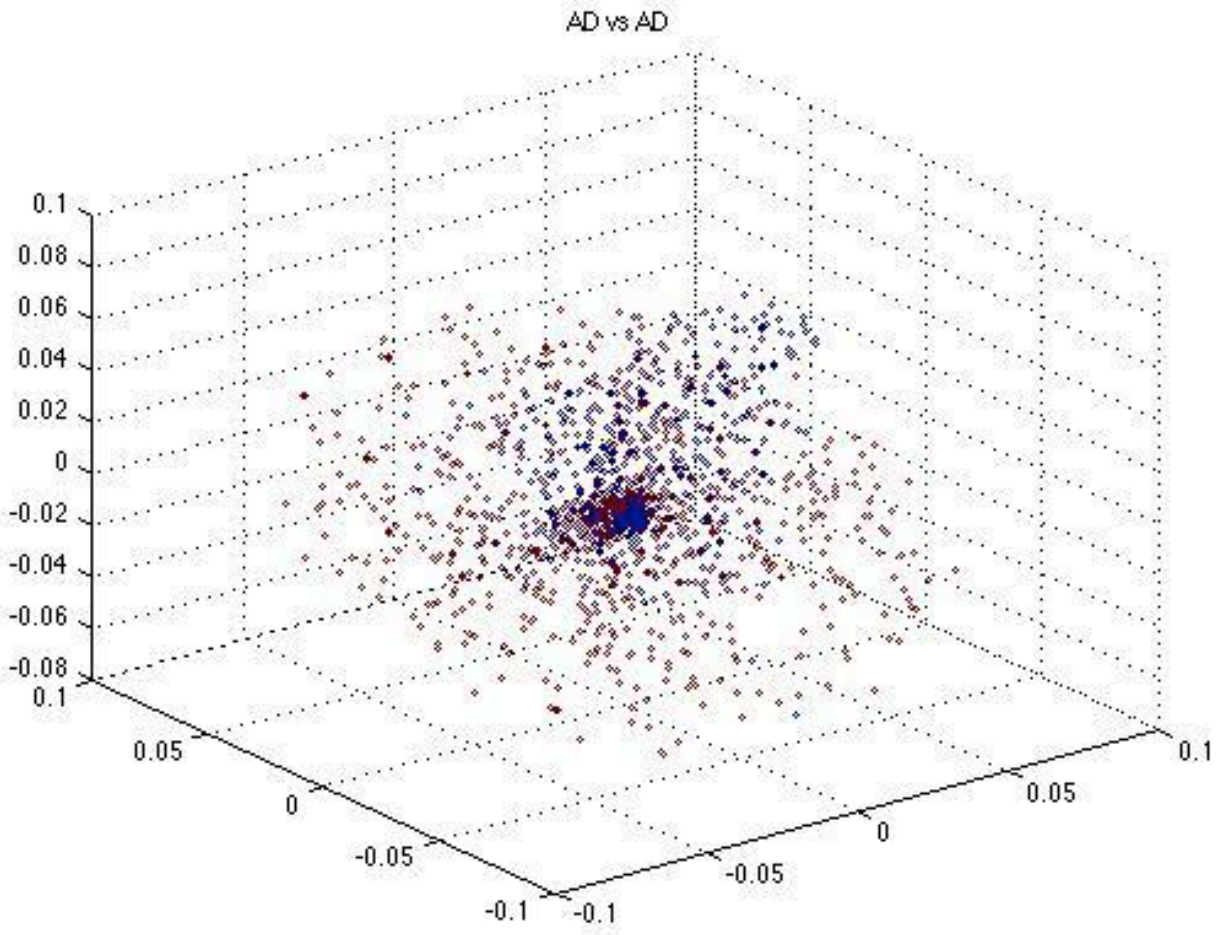
The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 2 patients' brains. The red points represent the data from the normal brain and the blue points correspond to the AD brain.

2 healthy brains



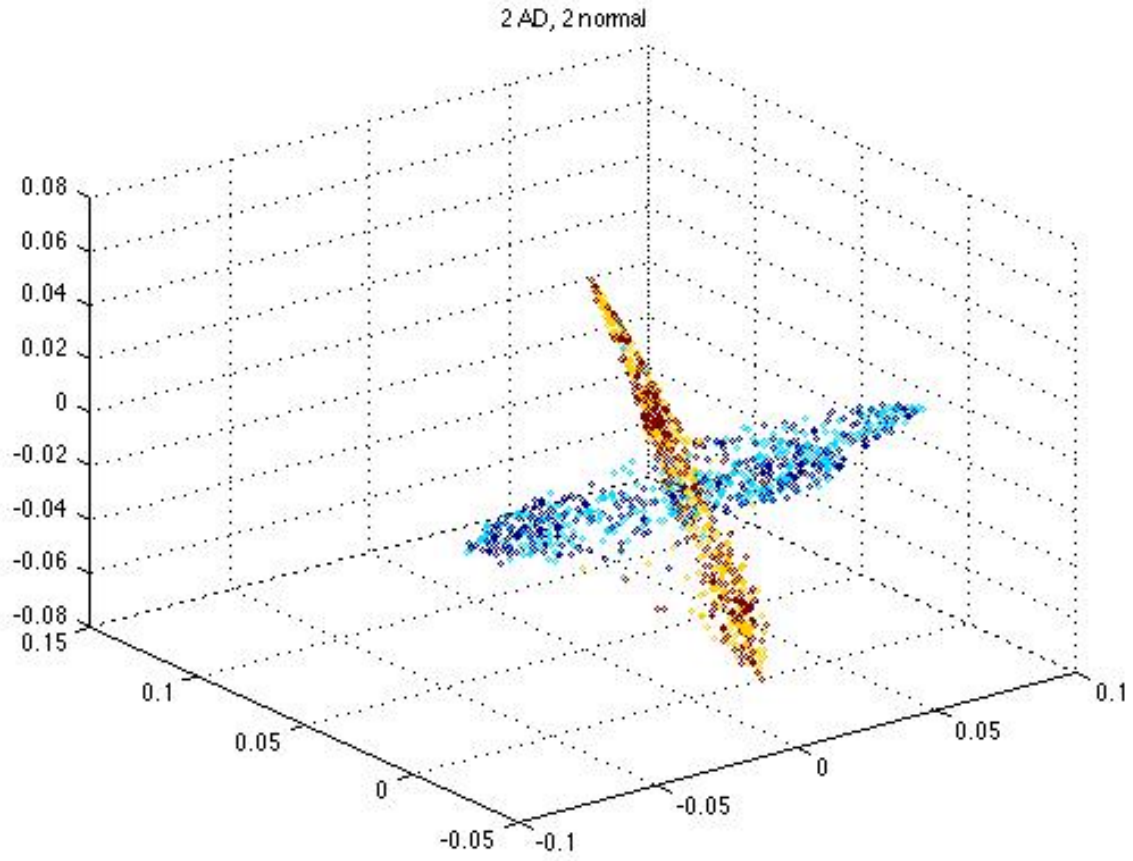
The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 2 healthy patients' brains.

2 Alzheimer's brains



The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 2 Alzheimer's patients' brains.

2 Alzheimer's and 2 healthy brains



The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 4 patients' brains. The red and yellow points represent the data from the AD brain and the light and dark blue points correspond to the normal brain.

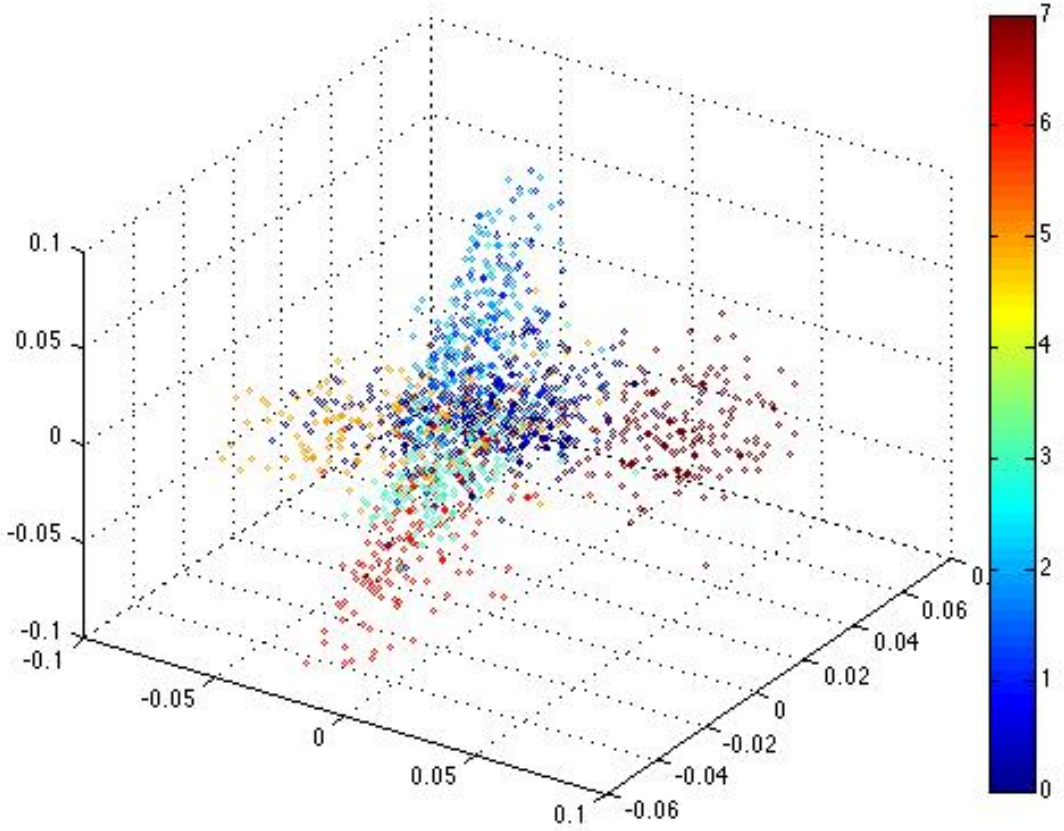
Earth Mover's Distance (EMD)

- Method to evaluate dissimilarity between 2 multi-dimensional distributions in some feature space where a distance measure between single features, the *ground distance,* is given. The EMD ``lifts'' this distance from individual features to full distributions.
- Intuitively, given 2 distributions, one can be seen as a mass of earth properly spread in space, the other as a collection of holes in that same space. Then, the EMD measures the least amount of work needed to fill the holes with earth.

The EMD has the following advantages

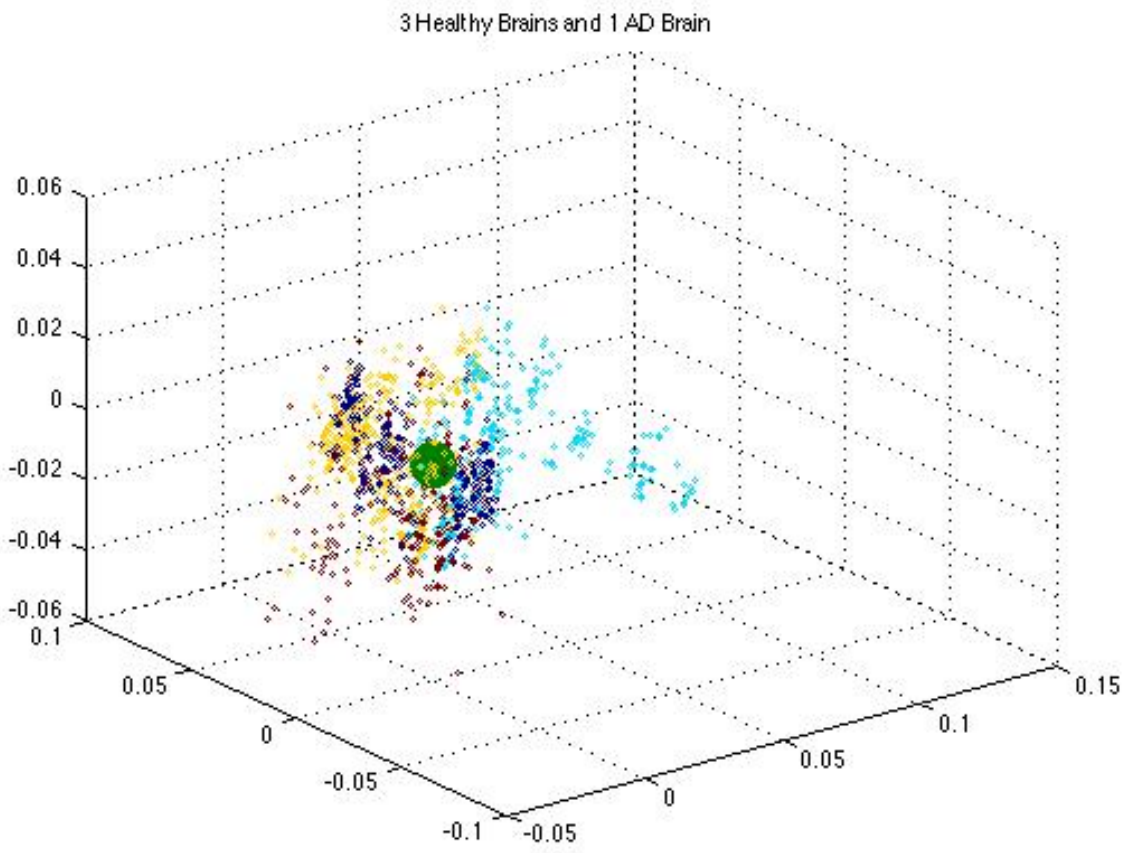
- Naturally extends the notion of a distance between single elements to that of a distance between sets, or distributions, of elements.
- Allows for partial matches in a very natural way. This is important, for instance, for image retrieval and to deal with occlusions and clutter.
- Is a true metric if the ground distance is metric and if the total weights of two signatures are equal. This allows endowing image spaces with a metric structure.

4 Alzheimer's and 4 healthy brains (using Earth Mover's Distance between histograms)



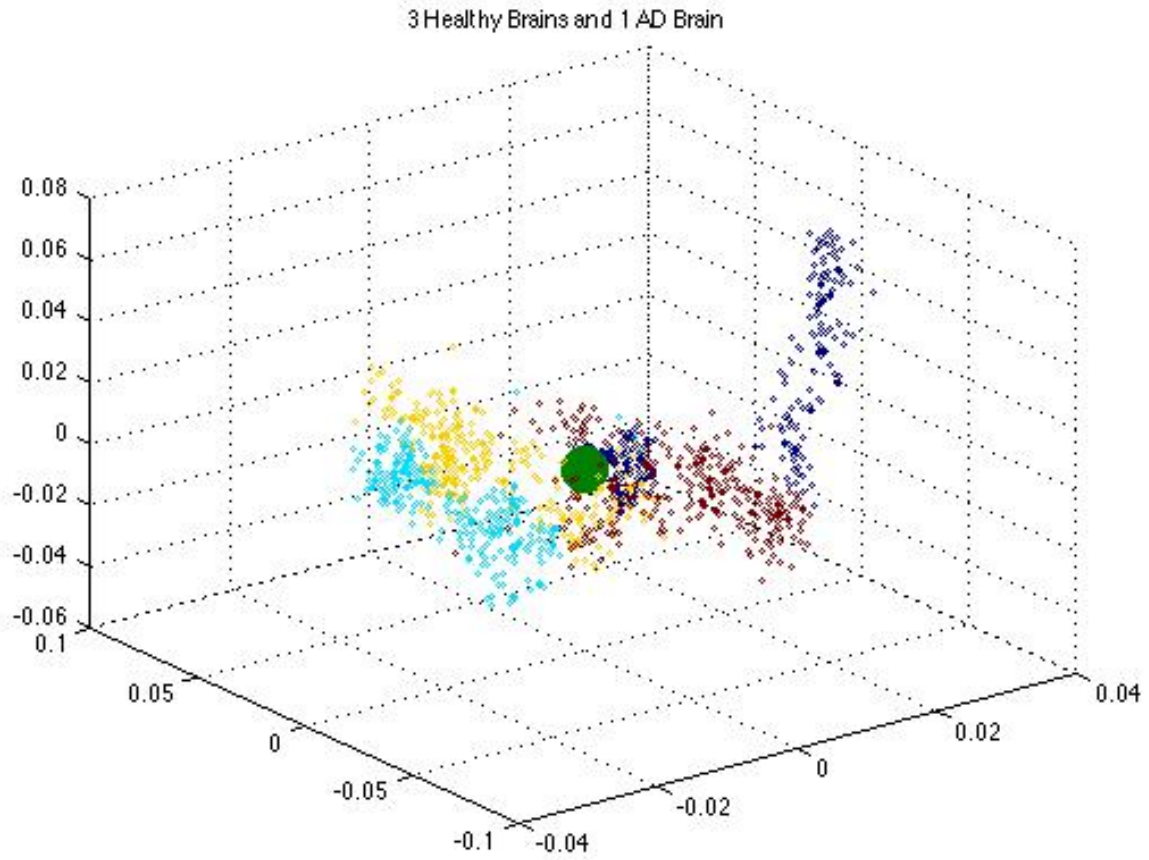
The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 8 patients' brains (4 healthy and 4 AD). This embedding was obtained using the Earth Mover's Distance between histograms.

1 Alzheimer's and 3 healthy brains



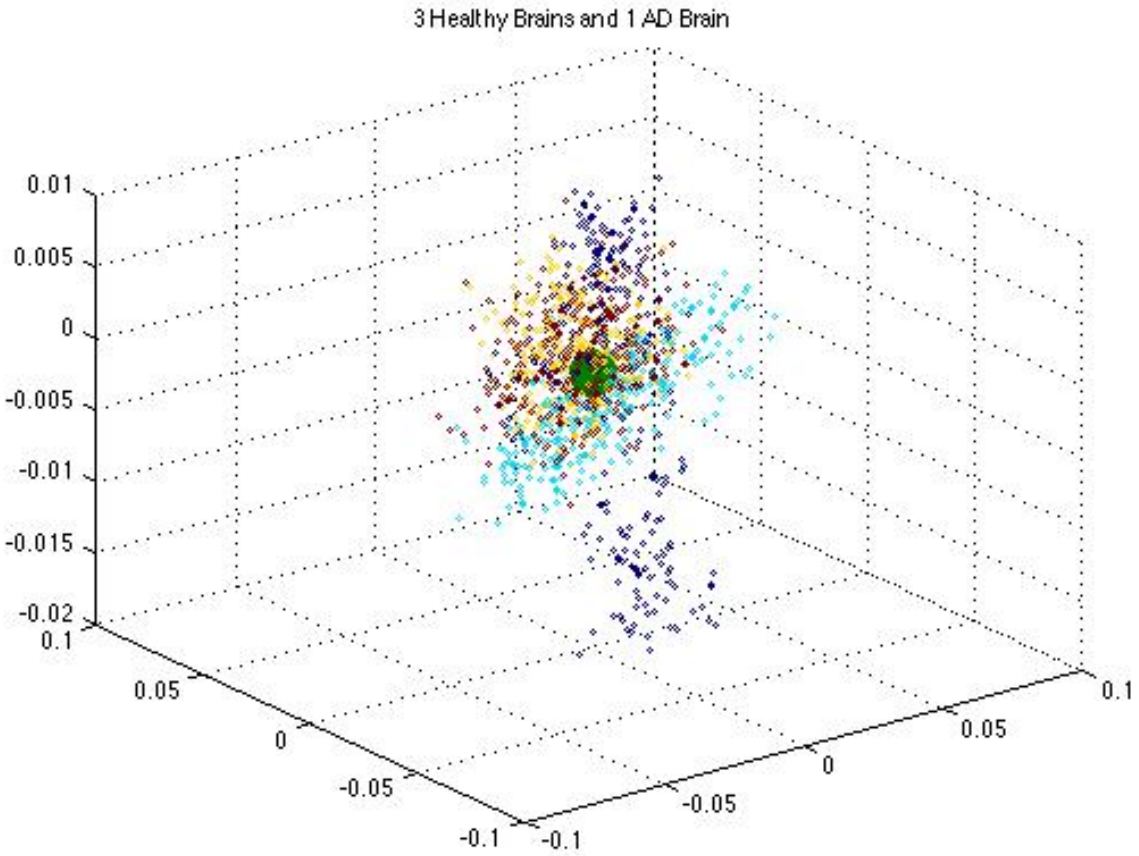
The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 4 patients' brains (3 healthy and 1 AD). The center of mass is represented by the large green dot and used to determine how spread out the points are, allowing us to choose which 3 eigenvectors to use for the embedding automatically.

Variance of AD points



The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 4 patients' brains (3 healthy and 1 AD). The center of mass is represented by the large green dot, and this embedding was chosen from all possible embeddings based on the maximum variance of AD points from the center of mass.

Variance Ratio



The figure above shows the NLICA embedding using 3 eigenvectors of the MRI data from 4 patients' brains (3 healthy and 1 AD). The center of mass is represented by the large green dot, and this embedding was chosen from all possible embeddings based on the maximum variance ratio of healthy points divided by AD points from the center of mass.

Discussion

- Using this nonlinear and local network approach, we were able to show a distinction between brains of patients with AD and brains of healthy patients.
- The combination of the local statistics and the Mahalanobis distance as well as using the Earth Mover's Distance between histograms is beneficial for such noisy MRI data
- This method is useful for nonlinear problems without existing definitive models

Future Work for Alzheimer's Disease prediction

- Find an automatic AD classification algorithm based on the obtained embeddings and use them to assist doctors in diagnosing AD
- Test the algorithm on more data and more patients
- Use the algorithm to distinguish among many different patients' brains

Acknowledgements

- Thank you to Professor Urszula Ledzewicz and the organizers of the conference for inviting me to give this talk
- Thank you to Thomas Strohmer, UC Davis, and to my collaborators at Yale, Ronald Coifman and Hitten Zaveri
- Thank you to NSF for support for this research (NSF-DTRA grant 1322393)

References

- Alzheimer's Association: Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia, 6 (2010), 158-194.
- R. Coifman, S. Lafon. Diffusion Maps. Appl. Comput. Harmon. Anal. 21 (2006), 5-30.
- D. Duncan, R. Talmon, H.P. Zaveri, R.R. Coifman. Identifying Preseizure State in Intracranial EEG Data Using Diffusion Kernels. Mathematical Biosciences and Engineering, 10 (2013), 579-90.
- D. Kushnir, A. Haddad, R.R. Coifman. Anisotropic diffusion on sub-manifolds with application to earth structure classification. Appl. Comp. Harm. Anal., 32 (2012), 280-294.
- A. Singer, R.R. Coifman. Non-linear independent component analysis with diffusion maps. Appl. Comp. Harmon. Anal. 25 (2008), 226-239.
- R. Talmon, D. Kushnir, R.R. Coifman, I. Cohen, S. Gannot. Parametrization of Linear Systems Using Diffusion Kernels. IEEE Transactions on Signal Processing, 60 (2012), 1-15,.
- W. Yang, R.L.M. Lui and X. Huang. Independent Component Analysis-Based Classification of Alzheimer's MRI data. J. Alzheimer's Dis. 24 (2011), 775-783.
- J. Ye et al. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC Neurology (2012), 12-46.