



Visualizing Alzheimer's disease progression in low dimensional manifolds

Kangwon Seo^a, Rong Pan^{b,*}, Dongjin Lee^b, Pradeep Thiyyagura^c, Kewei Chen^c, for the Alzheimer's Disease Neuroimaging Initiative¹

^a Department of Industrial and Manufacturing Systems Engineering and Department of Statistics, University of Missouri, USA

^b School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, USA

^c Banner Alzheimer's Institute, USA

ARTICLE INFO

Keywords:

Mathematics
Medical imaging
Visualization
AD progression
MR images
Nonlinear dimension reduction
Classification

ABSTRACT

While tomographic neuroimaging data is information rich, objective, and with high sensitivity in the study of brain diseases such as Alzheimer's disease (AD), its direct use in clinical practice and in regulated clinical trial (CT) still has many challenges. Taking CT as an example, unless the relevant policy and the perception of the primary outcome measures change, the need to construct univariate indices (out of the 3-D imaging data) to serve as CT's primary outcome measures will remain the focus of active research. More relevant to this current study, an overall global index that summarizes multiple complicated features from neuroimages should be developed in order to provide high diagnostic accuracy and sensitivity in tracking AD progression over time in clinical setting. Such index should also be practically intuitive and logically explainable to patients and their families. In this research, we propose a new visualization tool, derived from the manifold-based nonlinear dimension reduction of brain MRI features, to track AD progression over time. In specific, we investigate the locally linear embedding (LLE) method using a dataset from Alzheimer's Disease Neuroimaging Initiative (ADNI), which includes the longitudinal MRIs from 562 subjects. About 20% of them progressed to the next stage of dementia. Using only the baseline data of cognitively unimpaired (CU) and AD subjects, LLE reduces the feature dimension to two and a subject's AD progression path can be plotted in this low dimensional LLE feature space. In addition, the likelihood of being categorized to AD is indicated by color. This LLE map is a new data visualization tool that can assist in tracking AD progression over time.

1. Introduction

Alzheimer's Disease (AD), the most common cause of dementia, is accompanied with brain structure changes such as the whole or regional brain atrophy. Thus, in addition to the amyloid and tau related biomarkers, which are increasingly become informative (Jack et al. (2018)), diagnosing AD using neurodegenerative biomarkers such as the MRI-based volumetric measures is expected to provide supportive evidence to the conventional clinical diagnosis practice, which employs mainly cognitive tests and functional symptoms as well as the family medical history reviewed by physicians. While tomographic neuroimaging data is information rich, objective, and with high sensitivity in the study of brain diseases such as AD, its direct use in clinical practice and in regulated clinical trial (CT) still has many challenges. Taking CT as

an example, unless the relevant policy and the perception of the primary outcome measures change, the need to construct univariate indices (out of the 3-D imaging data) to serve as CT's primary outcome measure will remain the focus of many active researches (see, e.g., Chen et al. (2011); Chen et al. (2015); Fox et al. (2001); Hua et al. (2009); Fox et al. (2000); Thompson et al. (2004)). The voxel-based brain image analysis is an unbiased examination with intuitive brain image display. On the other hand, it has some serious concerns, including multiple comparisons of massive univariate outcomes (although there are ways to adjust the power of these comparisons), which is not feasible to serve as a CT outcome measure or as an explanation in clinical setting. Nevertheless, a number of studies have demonstrated that the neuroimaging-based univariate indices are accurate in diagnosis and sensitive to tracking changes over time (Jack et al. (2012)). The following list provides examples of

* Corresponding author.

E-mail address: rong.pan@asu.edu (R. Pan).

¹ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

<https://doi.org/10.1016/j.heliyon.2019.e02216>

Received 3 August 2018; Received in revised form 5 January 2019; Accepted 30 July 2019

2405-8440/© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

such indices.

- The volume of a pre-defined brain region (e.g., hippocampus volume) (Den Heijer et al. (2010))
- Glucose uptake in a pre-defined brain region (e.g., posterior cingulate) (Jagust et al. (2009); Landau et al. (2012))
- AD signature, an index defined by Mayo Clinics (Weston et al. (2016))
- The statistical region of interest (ROI) for tracking longitudinal changes (Chen et al. (2010))

Especially in the clinical setting, a single ROI-based index is a simplified version of disease marker that can be easily understood by patients and their families, but it may not be sensitive enough (as some of them are only derived from a local region (ROI), while others are introduced for different purposes, e.g., disease prognosis) for disease detection. One remedy is to summarize information over the whole brain (all voxels), not just a local region. This is a multivariate approach to defining a disease index.

One challenge from this approach is dimension reduction, in which information in high dimensional image data is represented in a low dimensional subspace by a set of principal attributes. The projected data in the low-dimensional space becomes more robust to noise and they can assist in building more stabilized diagnosis and prognosis models. In earlier studies, various linear dimension reduction techniques such as partial least square (PLS) (Phan et al. (2010)), principal component analysis (PCA) (Franke et al. (2010); Ayutyanont et al. (2010a,b); Khedher et al. (2015); Segovia et al. (2016)) on MR images have been studied and reported. However, it is recognized that the linear dimension reduction techniques may fail to convey proper latent information from high-dimensional space to low-dimensional space due to the complex non-linear patterns in dataset (Mwangi et al. (2014)). This is particularly true in handling neuroimaging data. Thus, nonlinear methods have been suggested by many researchers (e.g., Khajehnejad et al. (2017); Gerber et al. (2010); Hamm et al. (2010)). Gerber et al. (2010) investigated the applicability of manifold learning (nonlinear dimensionality reduction) on MR brain image data and Akhbardeh and Jacobs (2012) described the application of a novel scheme by combining wavelet transform and non-linear dimension reduction techniques to analyze breast MRI data.

Tracking AD progression over time has both clinical and research significance. Clinicians need to predict the progression of AD and medical researchers desire a good progression model so as to develop appropriate biomarkers or to design effective clinical trials. The standard method calculates the AD progression rate by examining patients on cognitions, functions and behaviors over time (Doody et al. (2010)). With the structural MRI data, the atrophies measured on some important ROIs, such as hippocampi, as well as on the whole brain have become a powerful biomarker for the identification of neurodegenerative stage and intensity in AD pathology (Vemuri and Jack (2010); Ayutyanont et al. (2010a,b); Ayutyanont et al. (2013)). In an early study, using voxel-based morphometry to map the structural change associated with conversion in MCI patients was discussed in Chetelat et al. (2005). More recently, structural MRI-based AD diagnosis approaches have been more extensively explored in the literature (see McEvoy et al. (2009); Magnin et al. (2009); Ewers et al. (2011); Lopez et al. (2011); Matoug et al. (2012); Lama et al. (2017); Long et al. (2017)). Most of them were using machine-learning tools such as support vector machine (SVM), import vector machine (IVM), regularized extreme learning machine (RELM), etc. In addition, longitudinal analysis and penalized regression methods have been applied on medical images to predict rapid-to-moderately-fast conversions from MCI to dementia (see Misra et al. (2009); McEvoy et al. (2011); Hinrichs et al. (2011); Moradi et al. (2015); Teipela et al. (2015); Korolev et al. (2016); Huang et al. (2017)).

In this paper, we develop a new visualization tool by using nonlinear dimension reduction of whole brain MRI features to track AD progression over time. A longitudinal dataset of high-dimensional features is represented in a two-dimensional space so that a patient's AD progression can

be easily traced and contrasted with the population model. Meanwhile, each lower dimensional data point also contains the information of higher dimensional features obtained from a classification model, and this leads to the minimum loss of prediction power. In order to effectively reduce dimensionality without losing AD classification accuracy, we apply a feature prescreening process on the data, and then apply a manifold-based dimension reduction technique, called locally linear embedding (LLE), on the reduced feature set. The data is mapped into a two-dimensional space and visualization of the data provides a convenient diagnosis for AD progression. As we apply LLE to a longitudinal dataset, the proposed method also allows us to track feature changes over time because the neighborhood structure of the reduced feature set is preserved on the LLE map. We use a machine-learning technique, support vector machine (SVM), to classify a subject being CU or AD. For each data point, the SVM model assigns a probability value of being classified to the AD category and colors the point based on this probability value. Therefore, we can visualize the AD progression of a patient by mapping out his/her LLE features on a two-dimensional chart over time. This visualization tool provides a direct depiction of the severity of AD and the speed of its progression, thus easy to be understood by patients.

In summary, the contributions of this paper are trifold – 1) a nonlinear dimension reduction technique is applied and it is better than linear dimension reduction in terms of capturing the internal structure of the high-dimensional feature space of brain images, thus improving the accuracy and sensitivity of SVM model in classifying AD and CU; 2) the visualization of these low-dimensional manifold features and the likelihood of AD, instead of brain images themselves, gives a more objective measure of AD progression over time; and 3) this new approach of feature fusion and visualization can help clinical prognosis and doctor-patient communication, and may further assist in AD research and drug development in the future.

The remaining of the paper is organized as follows: In Section 2.1, a brief description of dataset is provided. Then our data processing pipeline from feature prescreening to LLE representation is explained through Section 2.2 to 2.5. The validation of the proposed approach with the SVM model and the further development of the SVM model is provided in Sections 2.6 and 2.7. The application of this classification model on new observations is described in Section 2.7. In Section 3, the use of LLE map for AD progression tracking is explained and demonstrated on some patient data. Finally, Section 4 concludes the study.

2. Materials and methods

2.1. Brain MRI feature dataset

The brain MRI data used in this research are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. 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). For up-to-date information, see www.adni-info.org.

This dataset consists of multiple 1.5T MRI scans, corresponding to multiple visits over time for each of 761 subjects, that have been processed by FreeSurfer (version 4.4) (Reuter et al. (2012)). A total of 346 brain features, mostly the cortical thickness and volumes of various brain regions (e.g., volume of right superior temporal, cortical thickness average of left hippocampus) are stored in this dataset. The dataset also contains the QC output from manual inspection. Only the data with QC pass are used in our analysis, which give 2,402 scans of 562 subjects. The number of visits for a subject varies from 2 to 8 over the time period of years 2005–2011. Based on the diagnosis of the first visit (the baseline state), 177 cognitively unimpaired (CU), 277 mild cognitive impairment

(MCI), and 108 AD subjects are identified. Among them, 14 subjects progress from CU to MCI and 109 subjects progress from MCI to AD over the study's time period. The data analysis procedure applied on these 562 subjects is illustrated in Fig. 1.

2.2. Data preprocessing

The FreeSurfer image processing carries out the brain image registration based on the within-subject template. On top of that, in order to reduce the between-subject variability, each brain feature is divided by intracranial vault volume (ICV) of the corresponding subject so that the feature value can be interpreted as a fraction of brain volume, instead of a raw measurement value. After then, a unit normal scaling is applied to each feature, which is given as

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, \dots, 2402, \quad j = 1, \dots, 346. \quad (1)$$

where $\bar{x}_j$ and s_j are the average and the standard deviation, respectively, of the j^{th} feature.

2.3. Feature prescreening

The 346 original features may include some features that have little to no correlation with AD. Such features increase the magnitude of noise in the data and can hamper a successful dimension reduction. Therefore, we need to prescreen original features and select the ones that are moderately to highly correlated to AD. To do that, the correlation between a feature and the diagnosis result was evaluated by the intraclass correlation (IC), which is based on the following one-way random effect model:

$$z_{ijk} = \mu_j + b_{jk} + \varepsilon_{ijk}, \quad (2)$$

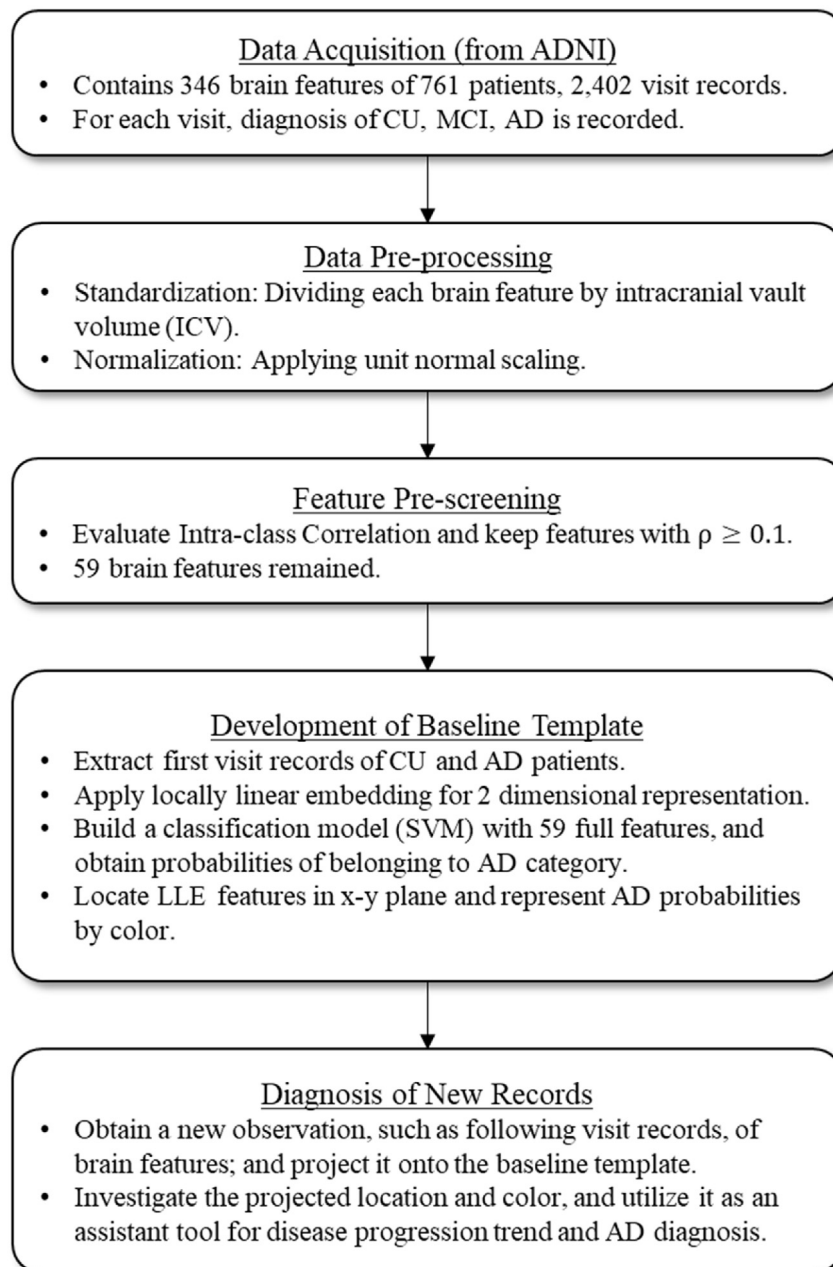


Fig. 1. An overview of data analysis process.

$$b_{jk} \sim N(0, \sigma_{b_j}^2), \quad (3)$$

$$\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon_j}^2), \quad (4)$$

where $k = \{CU, MCI, AD\}$ is the index for diagnosis group, z_{ijk} is the j^{th} feature value of the i^{th} instance which is diagnosed to the k^{th} category; μ_j is the overall mean of the j^{th} attribute; b_{jk} is the effect of group k ; and ε_{ijk} is the error term. The IC is defined to be

$$\rho_j = \frac{\sigma_{b_j}^2}{\sigma_{b_j}^2 + \sigma_{\varepsilon_j}^2}. \quad (5)$$

It is interpreted as the proportion of between-group variability in the total variability. Fig. 2 shows one example of a highly correlated feature (left) and another example of an almost independent feature (right) to AD diagnosis. We filter out the features that have ρ_j less than 0.1, and as a result, 59 brain features are retained for future analysis. A list of these 59 features is provided in Table 1.

2.4. Nonlinear dimension reduction by locally linear embedding

Previous studies have shown that the high-dimensional brain MRI features could be represented well in a low-dimensional space by using manifold-based dimension reduction techniques. Among others, Liu et al. (2013) applied locally linear embedding (LLE) to cross-sectional brain MRI feature data, and they showed that LLE features increased AD classification performance. Wolz et al. (2012) suggested a framework to incorporate subjects' meta-information into the manifold learning. Aljabar et al. (2011) proposed a combined representation of neonatal brain MR images from multiple features by using separate manifold learning steps.

In our study, we apply LLE to the prescreened longitudinal dataset explained in the previous section so as to illustrate the change of a brain over time. The basic idea of LLE is to force the neighborhood of an instance in the high-dimensional space to be preserved in the low-dimensional space with the same reconstruction weights (Roweis and Saul (2000); Saul and Roweis (2003)). Specifically, given a dataset $x_1, \dots, x_n \in \mathbb{R}^D$ in a high-dimensional space, we aim to find a low-dimensional representation $y_1, \dots, y_n \in \mathbb{R}^d$, where $d < D$, such that the neighborhood structure can be retained as much as possible. The LLE algorithm can be summarized by the following three steps:

1. For each data instance x_i , find the k nearest neighbors $x_{i1}, x_{i2}, \dots, x_{ik}$, by using Euclidean distance.
2. To find the optimal reconstruction weights in the original space, suppose x_i can be linearly reconstructed by its nearest neighbors, we let $\hat{x}_i = \sum_{j=1}^k w_{ij} x_{ij}$. Then, the reconstruction weights $w_{ij}, j = 1, \dots, k$, for each data point can be obtained by solving this optimization problem:

$$\min_w \left\| x_i - \sum_{j=1}^k w_{ij} x_{ij} \right\|^2, \quad \text{s.t.} \quad \sum_{j=1}^k w_{ij} = 1 \quad (6)$$

3. Finally, to find the new representation of the data in a reduced space, we again assume that y_i is a data point in the reduced space and it can be linearly reconstructed by its nearest neighbors, that is $\hat{y}_i = \sum_{j=1}^k w_{ij} y_{ij}$, where w_{ij} is the weights obtained from the previous step. Then, the LLE coordinates $Y = (y_1^T, y_2^T, \dots, y_n^T)^T$ can be obtained by solving

$$\min_Y \sum_{i=1}^n \left\| y_i - \sum_{j=1}^k w_{ij} y_{ij} \right\|^2, \quad \text{s.t.} \quad \sum_{i=1}^n y_i = 0, \quad Y^T Y = I \quad (7)$$

The objective function of Eq. (7) can be rewritten as $Y^T M Y$ where $M = (I - W)^T (I - W)$. Then, the solution is obtained by computing the eigenvectors of M that correspond to d small eigenvalues and using these eigenvectors to construct the columns of Y .

2.5. LLE representation

The dataset, as described in Section 2.1, includes repeated measurements from the same subject and these observations are typically correlated. To avoid any confounding effect of such correlation on LLE representation, we applied the LLE method on the first visit records of CU and AD subjects only. Different values of the number of nearest neighbors were tried, and each outcome was manually inspected to see whether two classes were reasonably separated in the two-dimensional space. As a result, $k = 16$ was chosen for the LLE representation of first visit records. Fig. 3a shows the LLE representation where the color of dots indicates the true label of diagnosis, either CU or AD. It is observed that these data points have a "V" shape, where CU subjects are more likely to be located on the left side of the shape and AD subjects on the right side. If a different dataset is used, LLE may produce a different orientation, e.g., the "V" shape may be upside down. To validate the consistency of this particular shape, we had repeated the analysis on the second visit and

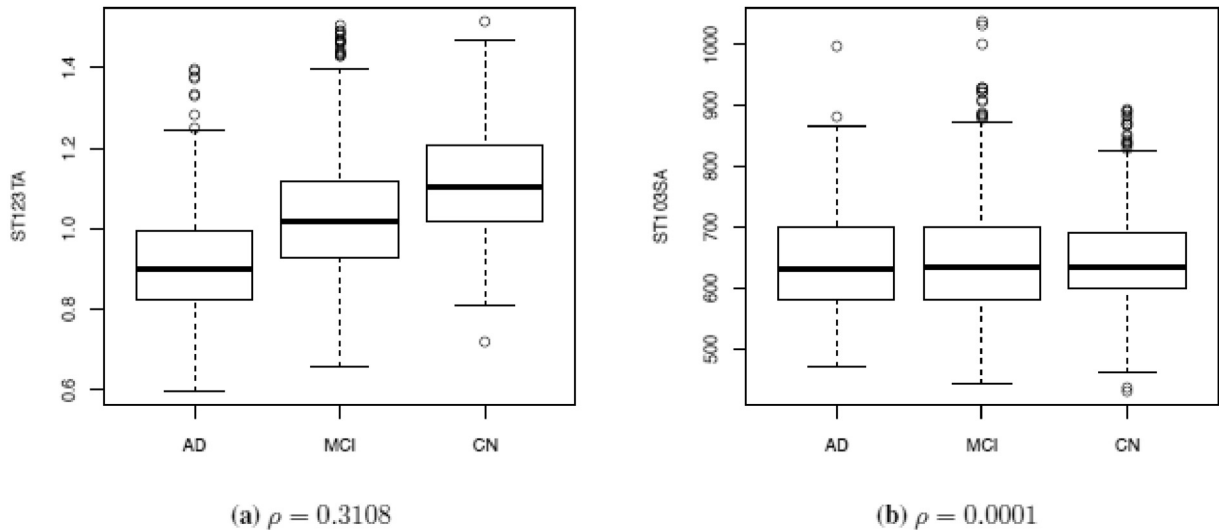


Fig. 2. Examples of brain features with (a) high correlation and (b) low correlation to AD diagnosis.

Table 1

A list of 59 prescreened features.

Feature name	Description
ST103CV	Volume (Cortical Parcellation) of RightParahippocampal
ST103TA	Cortical Thickness Average of RightParahippocampal
ST109CV	Volume (Cortical Parcellation) of RightPosteriorCingulate
ST111CV	Volume (Cortical Parcellation) of RightPrecuneus
ST111TA	Cortical Thickness Average of RightPrecuneus
ST115CV	Volume (Cortical Parcellation) of RightSuperiorFrontal
ST117CV	Volume (Cortical Parcellation) of RightSuperiorTemporal
ST117TA	Cortical Thickness Average of RightSuperiorTemporal
ST118CV	Volume (Cortical Parcellation) of RightSupramarginal
ST119TA	Cortical Thickness Average of RightTemporalPole
ST123CV	Volume (Cortical Parcellation) of RightUnknown
ST123TA	Cortical Thickness Average of RightUnknown
ST123TS	Cortical Thickness Standard Deviation of RightUnknown
ST129CV	Volume (Cortical Parcellation) of LeftInsula
ST12SV	Volume (WM Parcellation) of LeftAmygdala
ST130CV	Volume (Cortical Parcellation) of RightInsula
ST13CV	Volume (Cortical Parcellation) of LeftBanksts
ST13TA	Cortical Thickness Average of LeftBanksts
ST19SV	Volume (WM Parcellation) of LeftCerebralCortex
ST24CV	Volume (Cortical Parcellation) of LeftEntorhinal
ST24TA	Cortical Thickness Average of LeftEntorhinal
ST26CV	Volume (Cortical Parcellation) of LeftFusiform
ST26TA	Cortical Thickness Average of LeftFusiform
ST29SV	Volume (WM Parcellation) of LeftHippocampus
ST30SV	Volume (WM Parcellation) of LeftInferiorLateralVentricle
ST31CV	Volume (Cortical Parcellation) of LeftInferiorParietal
ST31TA	Cortical Thickness Average of LeftInferiorParietal
ST32CV	Volume (Cortical Parcellation) of LeftInferiorTemporal
ST32TA	Cortical Thickness Average of LeftInferiorTemporal
ST40CV	Volume (Cortical Parcellation) of LeftMiddleTemporal
ST40TA	Cortical Thickness Average of LeftMiddleTemporal
ST44CV	Volume (Cortical Parcellation) of LeftParahippocampal
ST44TA	Cortical Thickness Average of LeftParahippocampal
ST50CV	Volume (Cortical Parcellation) of LeftPosteriorCingulate
ST52CV	Volume (Cortical Parcellation) of LeftPrecuneus
ST52TA	Cortical Thickness Average of LeftPrecuneus
ST58CV	Volume (Cortical Parcellation) of LeftSuperiorTemporal
ST58TA	Cortical Thickness Average of LeftSuperiorTemporal
ST59CV	Volume (Cortical Parcellation) of LeftSupramarginal
ST60TA	Cortical Thickness Average of LeftTemporalPole
ST64CV	Volume (Cortical Parcellation) of LeftUnknown
ST64TA	Cortical Thickness Average of LeftUnknown
ST64TS	Cortical Thickness Standard Deviation of LeftUnknown
ST71SV	Volume (WM Parcellation) of RightAmygdala
ST72CV	Volume (Cortical Parcellation) of RightBanksts
ST72TA	Cortical Thickness Average of RightBanksts
ST78SV	Volume (WM Parcellation) of RightCerebralCortex
ST83CV	Volume (Cortical Parcellation) of RightEntorhinal
ST83TA	Cortical Thickness Average of RightEntorhinal
ST85CV	Volume (Cortical Parcellation) of RightFusiform
ST85TA	Cortical Thickness Average of RightFusiform
ST88SV	Volume (WM Parcellation) of RightHippocampus
ST89SV	Volume (WM Parcellation) of RightInferiorLateralVentricle
ST90CV	Volume (Cortical Parcellation) of RightInferiorParietal
ST90TA	Cortical Thickness Average of RightInferiorParietal
ST91CV	Volume (Cortical Parcellation) of RightInferiorTemporal
ST91TA	Cortical Thickness Average of RightInferiorTemporal
ST99CV	Volume (Cortical Parcellation) of RightMiddleTemporal
ST99TA	Cortical Thickness Average of RightMiddleTemporal

third visit data and found that this “V” shape of LLE features was indeed preserved with an appropriate tuning parameter k . Therefore, this shape reflects the innate nonlinear data structure of 59 features. Once a new observation of a subject is obtained, it can be projected onto this LLE map and the movement of LLE features summarizes the changes in the original features.

For the comparison purpose, we also constructed the two-dimensional representation based on the linear method, Principal Component Analysis (PCA). The PCA results are shown in Fig. 3b. One can see that these two representations are quite different. In PCA, the algorithm identifies two orthogonal directions with the largest global data variability in the high-dimensional space and projects the data to the

plane defined by these two directions. In contrast, LLE algorithm preserves the local structure of the original data points and tries to find the closest shape to this structure in a low-dimensional space. In specific, the relative distances among local neighborhoods in the original feature space are approximately saved on the LLE map. It suggests that the extents of AD progressions of multiple subjects located nearby are comparable by observing the location change on the LLE map. In addition, as can be seen from Fig. 3a, the “V” shape provides a relatively narrow path of AD progression along which data points can move, thus it enhances the AD progression visualization of longitudinal data. Note that the intrinsic dimensionality that covers 90% of data variability is found to be 9 for the LLE features and 21 for the PCA features, which also shows the efficiency of LLE features for dimension reduction.

2.6. Prediction power

Utilizing the new features derived from a dimension reduction algorithm for constructing predictive models has been discussed in the literature (see, e.g., Guerrero et al. (2011); Liu et al. (2013)). To investigate the prediction power of two dimensional LLE features, we first used the LLE features of the first visit records as training data to build a classification model, and then used the follow-up visit records as test data to evaluate the performance of this classification model. The Support Vector Machine (SVM) with Gaussian kernel function was chosen for the binary CU/AD classification. The model parameters were tuned by a 10-fold cross-validation. The same classification method was applied on the two dimensional PCA features too.

In Fig. 4, the receiver operating characteristic (ROC) curves of five classification models (with the LLE or PCA features only, with the LLE or PCA and original brain features, and with only original brain features) are plotted. First, it shows that there is not a significant difference in prediction power between the LLE and PCA features. For reference, the p -value of the statistical test (Robin et al. (2011)) to compare two ROC curves of LLE and PCA is 0.27, which indicates that those features perform similarly for classification. As these low dimensional representations are derived from the high dimensional original feature set, using these representations alone for subject classification is expected to lead to a loss of prediction power, as indicated by their ROCs. When combined with original features, the LLE method, represented by “LLE + full features” in Fig. 4, starts to demonstrate better performance in AD classification compared to the “PCA + full features”. The advantage of LLE comes from its nonlinear manifold embedding capability, which preserves the intrinsic data structure within the original dataset. However, as shown in Fig. 4 where a majority of the blue solid curve is indistinguishable from the grey solid curve, the improvement of the “LLE + full features” curve is found to be minimal as compared with the ROC curve produced by the 59 original features (i.e., full features), which suggests it is sufficient to use only original features for classification on the LLE map.

We complement the LLE map by taking advantage of the prediction power of the original 59 features. That is, the probability of AD is obtained from the SVM model with the original 59 features and it is represented by color (from blue to red) for each data point on the map. We call this complemented map the baseline template, as shown in Fig. 5, where the red dots indicate the more AD likely patients and the blue dots the more CU likely. This template provides a population model for the subjects under study and it plays a critical role in contrasting an individual subject's AD progression to the population pattern.

2.7. Projection of new observations

New observations, such as the follow-up visit records of the subjects who are included in the baseline template or the subjects who are diagnosed with MCI at their first visit, can be projected onto the baseline template (see Fig. 6) by the following procedure:

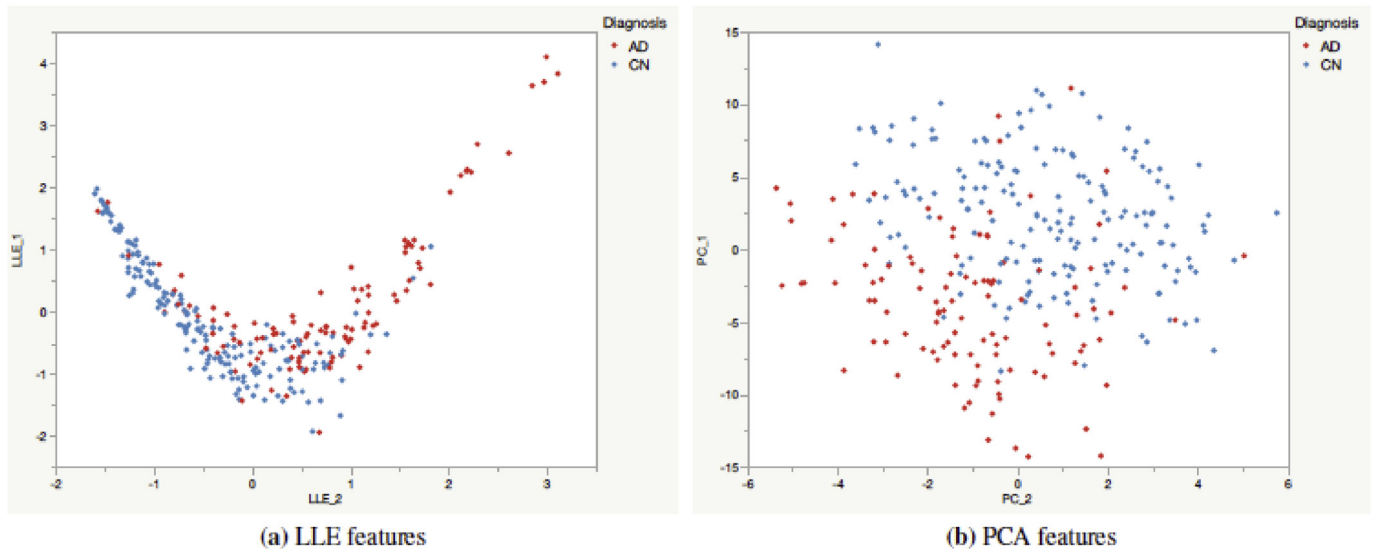


Fig. 3. (a)LLE and (b)PCA representations of first visit records of CU and AD patients.

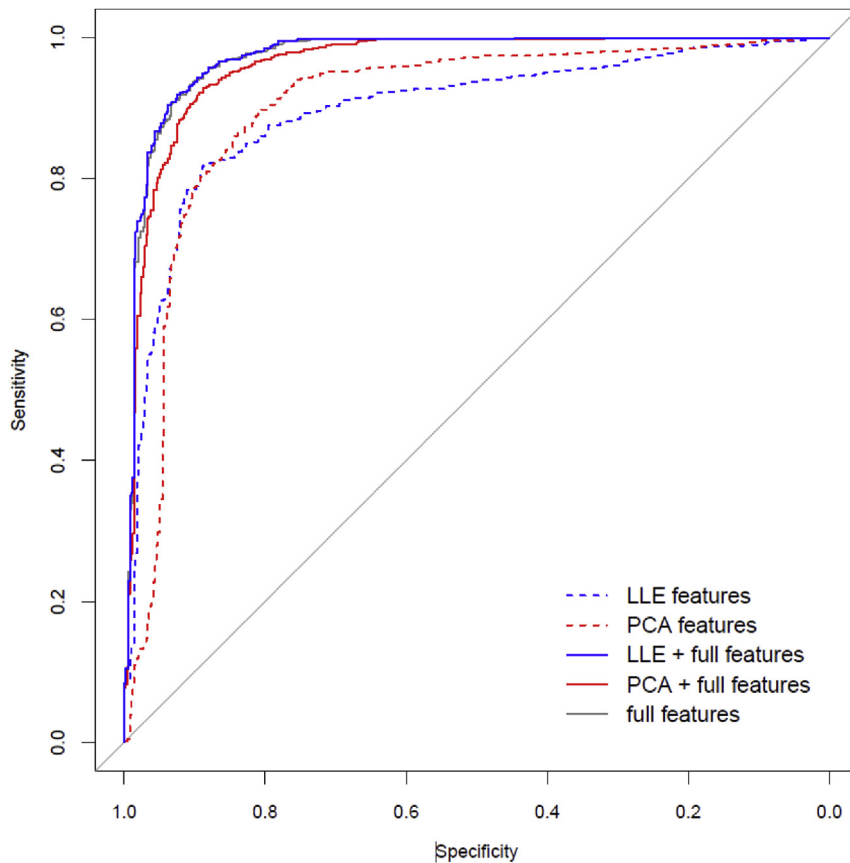


Fig. 4. Comparison of ROC curves from binary (CU/AD) classification models using 1) LLE features (blue dashed), 2) PCA features (red dashed), 3) LLE and original features (blue solid), 4) PCA and original features (red solid), and 5) original features only (grey solid).

1. Among the instances included in the baseline template, find the k nearest neighbors ($k = 16$ for the baseline template in Fig. 5) of the new data point x_0 in the original 59-feature space.
2. Compute the linear reconstruction weight w_{0j} for each neighbor by Eq. (6).
3. Compute two LLE coordinates of the new data point by linearly combining LLE coordinates of the nearest neighbors using the same weights. That is

$$y_0 = \sum_{j=1}^k w_{0j} y_{0j} \quad (8)$$

4. Compute the probability of the new observation being classified to the AD category, p_0 , by either 1) applying the SVM model built for the baseline template or 2) a linear combination of the probabilities of k nearest neighbors, p_{0j} , using the same weights; that is

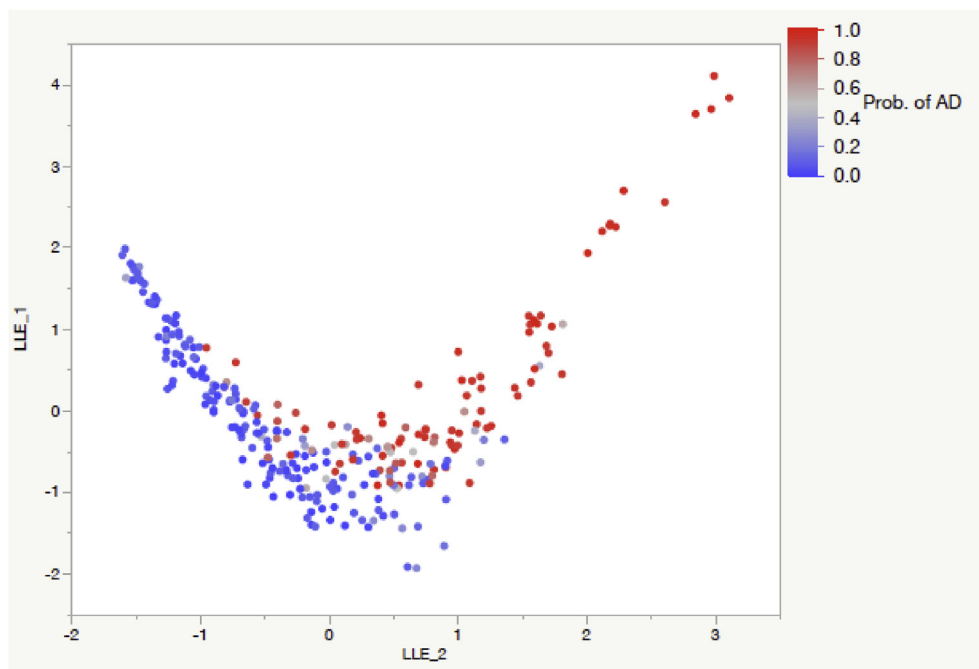


Fig. 5. Baseline template with probability of belonging to AD category.

$$p_0 = \sum_{j=1}^k w_{0j} p_{0j} \quad (9)$$

The first approach is used in this paper because it is more accurate, as the full feature information is applied to the SVM model. However, the second approach is useful when the classification model built for the baseline template is missing or too complicated to be computationally efficient.

3. Results

Fig. 6 gives the visualization of AD diagnoses of 6 subjects over 6 or 7 visits. All of these subjects are identified as MCIs on their first visit, but the paths they took to progress to the AD state are quite different. For example, Subjects 108 and 214 stays in the non-AD region at the bottom of the “V” shape for a longer period before moving to the AD region; while Subjects 269 and 631 move quickly to the AD region and even accelerate the speed of moving to the far right of the “V” shape at later visits. Therefore, comparing to the population under study, the mental health of Subject 269 and Subject 631 are deteriorating faster and, most likely, this trend will become worse in future.

Each of 562 subjects’ progression path can be depicted in the LLE AD progression map. We observe the subject started at some location on the map gradually moving to the right side over subsequent visits. Also, we can observe the changing AD probability for each subject from the colors of sequential points. The main diagnosis characteristics that we may obtain from this LLE map are as follows:

- The current location: Where is the patient's mental health located on the baseline template? Is it on the left arm of “V” or the right arm of “V”?
- The probability of AD: What does the classification model evaluate the current health status? How likely is the patient to be an AD patient?
- The direction of AD progression: Does the patient's health state tend to stay in the same region as from the previous visits? In what direction does it proceed?

- The progression rate: How fast does the patient's health state move to the AD region? Is there a big change of health state between two consecutive visits?

This visualization tool is able to depict AD progression more clearly than merely examining the original 3D brain images. For example, identifying the change of the size of ventricle is one of the traditional ways of the image-based AD diagnosis, but sizing ventricle from images is not a reliable method. To emphasize this point, the horizontal slices of MRI scans corresponding to the sequential visit records of a subject in Fig. 6c (RID = 214) are shown in Fig. 7. The changes in ventricle area from these images are difficult to detect by human eyes, which may cause a failure in identifying and quantifying the disease's progression if the radiologist only reads these images. On the other hand, the changes in Fig. 6c are clear, and they provide an objective assessment of disease progression that cannot be overlooked.

4. Discussion

In this study, we propose a new visualization tool for tracking AD progression over time. The brain MRI features are represented in a two-dimensional LLE feature map. As the classification accuracy for CU and AD with 2 LLE features is only about 83%, which is inferior to the one with full features, 92%, the probability of belonging to the AD category will be computed from full features and depicted by a corresponding color, which can be thought of as the third dimension. The proposed method is capable of providing a convenient and intuitive AD diagnosis result, where the progression of the disease can be monitored. Besides, a SVM classifier trained with the LLE features and the full 59 prescreened features can achieve very high classification accuracy. Along with other clinical tests, we believe this new approach is promising in extracting information from longitudinal brain image data for better tracking the AD progression over time.

A particular observation in this study is that the application of manifold-based dimension reduction technique on MRI brain image data does not ensure its outperformance, in terms of the classification of CU and AD patients, over the linear dimension reduction technique by itself.

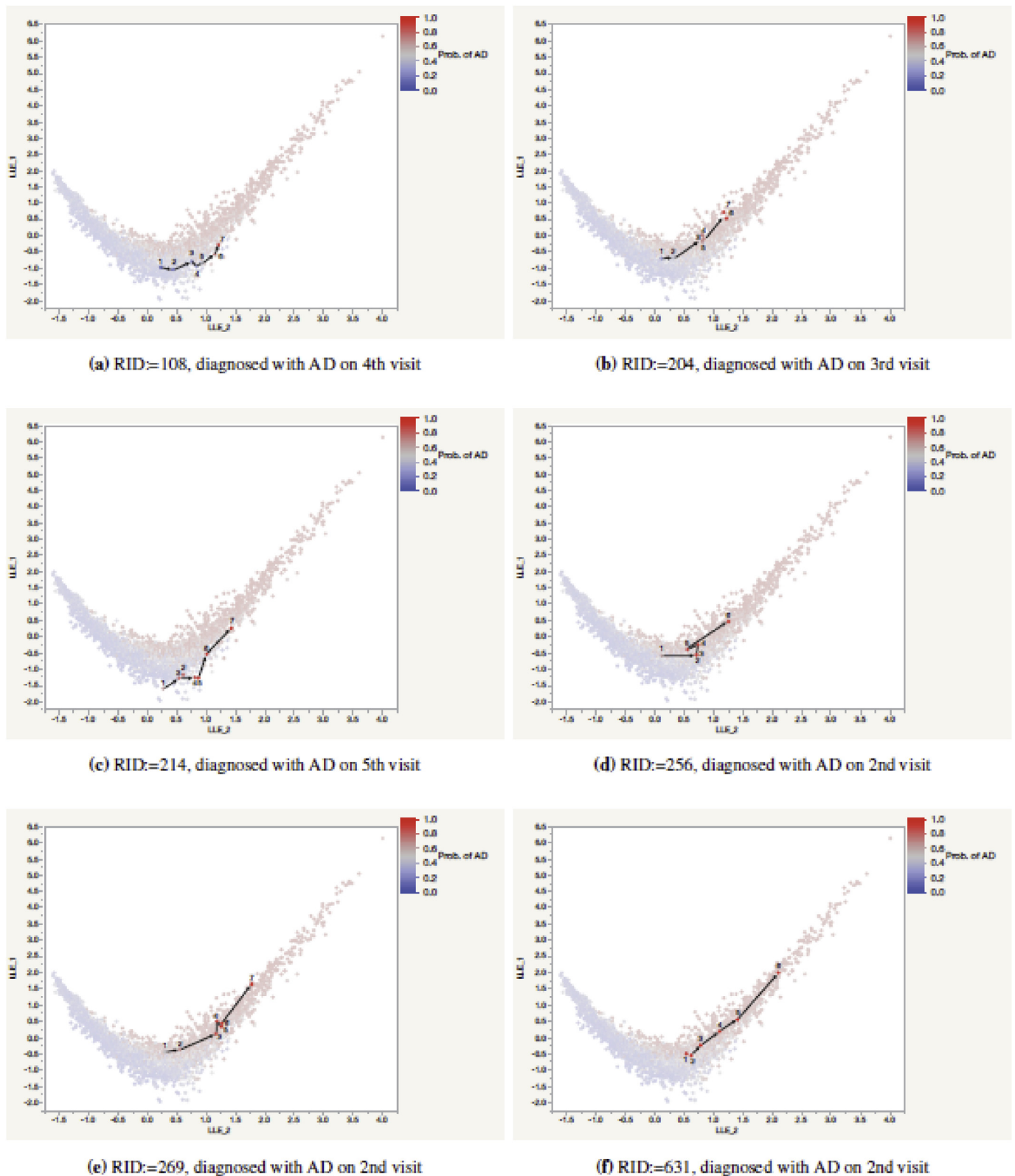


Fig. 6. Examples of AD progression paths. A patient's AD progression path was represented by dots with visiting sequence and arrows. Each patient in these cases was diagnosed with MCI on the first visit and progressed to AD on following visits. Colors of dots in background were dimmed out for clarification of a progression path. RID is the subject ID.

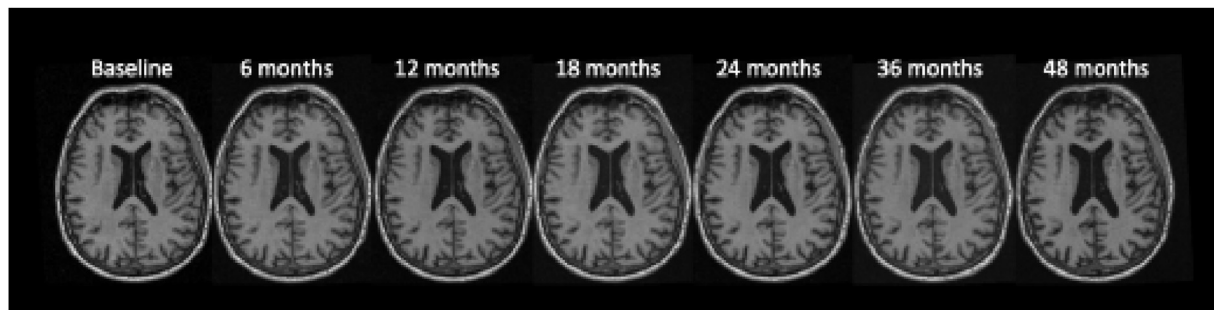


Fig. 7. Horizontal slices of MRI scans over a 4-year follow-up period for a subject with RID = 214. The changes in the size of ventricular and subarachnoid spaces are subtle.

As shown in Section 2.6, the two LLE features alone do not necessarily perform better than the PCA features on classifying CU and AD patients. In fact, Van Der Maaten et al. (2009) describe that PCA is likely to perform better than non-linear method for a real-world dataset. A possible reason could be the violation of the fundamental assumption of LLE algorithm, i.e. the local linearity, in the real-world dataset. However, as shown in Section 2.7, because LLE features are able to preserve the local structure of a dataset, they complement the original data in a classification algorithm. Thus, by combining LLE features with original features we can achieve a better classification performance over the linear dimension reduction technique. The multicollinearity between the original features and the PCA features, which are linear combinations of original features, also contributes to the inferior performance of PCA.

In this paper, we used the dataset containing the brain features extracted from the FreeSurfer processing, which requires the knowledge of 3D modeling for brain images and it is not always guaranteed that a new observation could be processed by the same process due to the software availability and the variability in quality control. In future research, we will investigate applications of other dimension reduction methods and machine learning techniques, such as t-SNE (Van Der Maaten and Hinton, 2008), ISOMAP (Tenenbaum et al., 2000), random forests and deep learning, in order to explicitly reveal complex latent patterns in voxel-based brain image data and to build robust models. If we can directly utilize voxel data without brain feature extraction, it will provide a simpler and unified method for the image-based AD diagnosis and prognosis.

Declarations

Author contribution statement

Kangwon Seo, Rong Pan, Kewei Che: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Dongjin Lee: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Pradeep Thiyyagura: Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the 2016 Pilot grant from Arizona Alzheimer's Consortium, USA and Banner Alzheimer's foundation.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study has been deposited at the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, <http://adni.loni.usc.edu>.

References

- Akhbardeh, A., Jacobs, M.A., 2012. Comparative analysis of nonlinear dimensionality reduction techniques for breast mri segmentation. *Med. Phys.* 39 (4), 2275–2289.
- Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S.J., Rutherford, M.A., Edwards, A.D., Rueckert, D., 2011. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. *IEEE Trans. Med. Imaging* 30 (12), 2072–2086.
- Ayutyanont, N., Chen, K., et al., 2010a. Twelve-month whole-brain atrophy rates and estimated power to detect Alzheimer-slowing treatment effects in multicenter trials using iterative principal component analysis: preliminary findings from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia. J. Alzheimer's Assoc.* 6 (4), S303.
- Ayutyanont, N., Chen, K., et al., 2013. Whole brain atrophy and sample size estimate via iterative principal component analysis for twelve-month Alzheimer's disease trials. *Neurosci. Biomed. Eng.* 1 (1).
- Ayutyanont, N., Chen, K., Reschke, C., Langbaum, J.B., Fleisher, A., Kong, L., Reiman, E.M., 2010b. Twelve-month whole-brain atrophy rates and estimated power to detect Alzheimer-slowing treatment effects in multicenter trials using iterative principal component analysis: preliminary findings from the alzheimers disease neuroimaging initiative. *Alzheimer's Dementia* 6 (4), S303.
- Chen, K., Ayutyanont, N., Langbaum, J.B., Fleisher, A.S., Reschke, C., Lee, W., et al., 2011. Characterizing Alzheimer's disease using a hypometabolic convergence index. *Neuroimage* 56 (1), 52–60.
- Chen, K., Langbaum, J.B.S., Fleisher, A.S., et al., 2010. Twelve-month metabolic declines in probable Alzheimer's disease and amnesic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's disease neuroimaging initiative. *Neuroimage* 51 (2), 654–664.
- Chen, K., Roontiva, A., Thiyyagura, P., Lee, W., Liu, X., Ayutyanont, N., et al., 2015. Improved power for characterizing longitudinal amyloid- β pet changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J. Nucl. Med.* 56 (4), 560–566.
- Chetelat, G., et al., 2005. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. *Neuroimage* 27, 934–946.
- Den Heijer, T., van der Lijn, F., Koudstaal, P.J., Hofman, A., van der Lugt, A., Krestin, G.P., Niessen, W.J., Breteler, M.M.B., 2010. A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain* 133 (4), 1163–1172.
- Doody, R.S., Pavlik, V., Massman, P., Rountree, S., Darby, E., Chan, W., 2010. Predicting progression of Alzheimer's disease. *Alzheimer's Res. Ther.* 2, 2.
- Ewers, M., Sperling, R.A., Klunk, W.E., Weiner, M.W., Hampel, H., 2011. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci.* 34 (8).
- Fox, N.C., Cousens, S., Scallin, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. *Arch. Neurol.* 57 (3), 339–344.
- Fox, N.C., Crum, W.R., Scallin, R.I., Stevens, J.M., Janssen, J.C., Rossor, M.N., 2001. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *The Lancet* 358 (9277), 201–205.
- Franke, K., Ziegler, G., Klippel, S., Gaser, C., Alzheimer's Disease Neuroimaging Initiative, 2010. Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: exploring the influence of various parameters. *Neuroimage* 50 (3), 883–892.

- Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R., Alzheimer's Disease Neuroimaging Initiative, 2010. Manifold modeling for brain population analysis. *Med. Image Anal.* 14 (5), 643–653.
- Guerrero, R., Wolz, R., Rueckert, D., 2011. Laplacian eigenmaps manifold learning for landmark localization in brain mr images. In: *International Conference on Medical Image Computing and Computer Assisted Intervention* (Pp. 566–573).
- Hamm, J., Ye, D.H., Verma, R., Davatzikos, C., 2010. Gram: a framework for geodesic registration on anatomical manifolds. *Med. Image Anal.* 14 (5), 633–642.
- Hinrichs, C., Singh, V., Xu, G.F., Johnson, S.C., Alzheimer's Disease Neuroimaging Initiative, 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 55, 574–589.
- Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Chou, Y.-Y., Ho, A.J., et al., 2009. Optimizing power to track brain degeneration in alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ANDI study of 515 subjects. *Neuroimage* 48 (4), 668–681.
- Huang, M., Yang, W., Feng, Q., Chen, W., the Alzheimer's Disease Neuroimaging Initiative, 2017. Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Nature, Sci. Rep.* 7, 39880.
- Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., et al., 2018. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dementia* 14 (4), 535–562.
- Jack, C.R., Vemuri, P., Wiste, H.J., et al., 2012. Shapes of the trajectories of five major biomarkers of alzheimer's disease. *Arch. Neurol.* 69 (7), 856–867.
- Jagust, W.J., Landau, S.M., Shaw, L.M., et al., 2009. Relationships between biomarkers in aging and dementia. *Neurology* 73 (15), 1193–1199.
- Khajehnejad, M., Saatlou, F.H., Mohammadzade, H., 2017. Alzheimer's disease early diagnosis using manifold-based semi-supervised learning. *Brain Sci.* 7, 109.
- Khedher, L., Ramirez, J., Grizz, J.M., Brahim, A., Segovia, F., Alzheimer's Disease Neuroimaging Initiative, 2015. Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented mri images. *Neurocomputing* 151, 139–150.
- Korolev, I.O., Symonds, L.L., Bozoki, A.C., 2016. Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. *PLoS One* 11, e0138866.
- Lama, R.K., Gwak, J., Park, J.-S., Lee, S.-W., 2017. Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features. *J. Healthcare Eng.* 2017, 5485080, 11 pages.
- Landau, S.M., Mintun, M.A., Joshi, A.D., et al., 2012. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann. Neurol.* 72 (4), 578–586.
- Liu, X., Tosun, D., Weiner, M.W., Schuff, N., Initiative, A.D.N., et al., 2013. Locally linear embedding (lle) for mri based alzheimer's disease classification. *Neuroimage* 83, 148–157.
- Long, X., Chen, L., Jiang, C., Zhang, L., Alzheimer's Disease Neuroimaging Initiative, 2017. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS One* 12 (3), e0173372.
- Lopez, M., et al., 2011. Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. *Neurocomputing* 74, 1260–1270.
- Magnin, B., et al., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* 51, 73–83.
- Matoug, S., Abdel-Dayem, A., Passi, K., Gross, W., Alqarni, M., 2012. Predicting Alzheimer's disease by classifying 3D brain MRI images using SVM and other well defined classifiers. *J. Phys. Conf. Ser.* 341, 012019.
- McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., the Alzheimer's Disease Neuroimaging Initiative, 2009. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology* 251 (1).
- McEvoy, L.K., et al., 2011. Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology* 259, 834–843.
- Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* 44, 1415–1422.
- Moradi, E., et al., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104, 398–412.
- Mwangi, B., Tian, T.S., Soares, J.C., 2014. A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12 (2), 229–244.
- Phan, T.G., Chen, J., Donnan, G., Srikanth, V., Wood, A., Reutens, D.C., 2010. Development of a new tool to correlate stroke outcome with infarct topography: a proof-of-concept study. *Neuroimage* 49 (1), 127–133.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61 (4), 1402–1418.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Muller, M., 2011. Proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinf.* 12 (1), 77.
- Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290 (5500), 2323–2326.
- Saul, L.K., Roweis, S.T., 2003. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res.* 4, 119–155.
- Segovia, F., Garca-Prez, M., Grizz, J.M., Ramirez, J., Martinez-Murcia, F.J., 2016. Assisting the diagnosis of neurodegenerative disorders using principal component analysis and tensorflow. In: *International Conference on European Transnational Education*, pp. 43–52.
- Teipela, S.J., Kurth, J., Krause, B., Grothe, M.J., Alzheimer's Disease Neuroimaging Initiative, 2015. The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment — beyond classical regression. *Neuroimage: Clin.* 8, 583–593.
- Tenenbaum, J.B., De Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290 (5500), 2319–2323.
- Thompson, P.M., Hayashi, K.M., De Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., et al., 2004. Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage* 22 (4), 1754–1766.
- Van Der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9 (Nov), 2579–2605.
- Van Der Maaten, L., Postma, E., Van den Herik, J., 2009. Dimensionality reduction: a comparative. *J. Mach. Learn. Res.* 10, 66–71.
- Vemuri, P., Jack, C.R., 2010. Role of structural MRI in Alzheimer's disease. *Alzheimer's Res. Ther.* 2, 23.
- Weston, P.S.J., Nicholas, J.M., Lehmann, M., et al., 2016. Presymptomatic cortical thinning in familial Alzheimer disease: a longitudinal MRI study. *Neurology* 87 (19), 2050–2057.
- Wolz, R., Aljabar, P., Hajnal, J.V., Lotjonen, J., Rueckert, D., Initiative, A.D.N., et al., 2012. Nonlinear dimensionality reduction combining mr imaging with non-imaging information. *Med. Image Anal.* 16 (4), 819–830.