

Early diagnosis of Alzheimer's disease on ADNI data using novel longitudinal score based on functional principal component analysis

Haolun Shi,^{a,†} Da Ma^{b,†}, Yunlong Nie,^a Mirza Faisal Beg^{b,‡}, Jian Pei,^{a,c,‡} Jiguo Cao^b,^{a,c,‡,*} and The Alzheimer's Disease Neuroimaging Initiative[§]

^aSimon Fraser University, Department of Statistics and Actuarial Science, Burnaby, BC, Canada

^bSimon Fraser University, School of Engineering Science, Burnaby, BC, Canada

^cSimon Fraser University, School of Computing Science, Burnaby, BC, Canada

Abstract

Methods: Alzheimer's disease (AD) is a worldwide prevalent age-related neurodegenerative disease with no available cure yet. Early prognosis is therefore crucial for planning proper clinical intervention. It is especially true for people diagnosed with mild cognitive impairment, to whom the prediction of whether and when the future disease onset would happen is particularly valuable. However, such prognostic prediction has been proven to be challenging, and previous studies have only achieved limited success.

Approach: In this study, we seek to extract the principal component of the longitudinal disease progression trajectory in the early stage of AD, measured as the magnetic resonance imaging (MRI)-derived structural volume, to predict the onset of AD for mild cognitive impaired patients two years ahead.

Results: Cross-validation results of LASSO regression using the longitudinal functional principal component (FPC) features show significant improved predictive power compared to training using the baseline volume 12 months before AD conversion [area under the receiver operating characteristic curve (AUC) of 0.802 versus 0.732] and 24 months before AD conversion (AUC of 0.816 versus 0.717).

Conclusions: We present a framework using the FPCA to extract features from MRI-derived information collected from multiple timepoints. The results of our study demonstrate the advantageous predictive power of the population-based longitudinal features to predict the disease onset compared with using only cross-sectional data-based on volumetric features extracted from a single timepoint, demonstrating the improved prediction power using FPC-derived longitudinal features.

© 2021 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.8.2.024502](https://doi.org/10.1117/1.JMI.8.2.024502)]

Keywords: Alzheimer's disease; dementia of the Alzheimer type; early prediction; longitudinal; functional principal component analysis; Alzheimer's disease neuroimaging initiative.

Paper 20108RR received May 22, 2020; accepted for publication Mar. 12, 2021; published online Apr. 21, 2021.

1 Introduction

Alzheimer's disease (AD) is a type of dementia associated with age-related neurodegeneration and cognitive decline. AD affects a large proportion of the population, anticipated to worsen with

*Address all correspondence to Jiguo Cao, jiguo_cao@sfu.ca

[†]Joint first author

[‡]Joint senior author

[§]Data used in preparation of this article were partly obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://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.

increasing societal longevity. Predicting AD onset at an early stage is important for improving patients care and developing therapeutic interventions. Medical imaging techniques, such as magnetic resonance imaging (MRI) and fluorodeoxyglucose positron emission tomography (PET), have been playing a crucial role in the diagnosis of AD before the manifest of clinical symptoms.¹

Depending on the severity of the symptom, the diagnoses of AD are categorized into three stages: cognitive normal (CN), mild cognitive impairment (MCI), and finally being diagnosed as AD. Subjects diagnosed as MCI are regarded as the at-risk group, and the prediction of their disease onset is important for planning of the patient care, early intervention, and potential treatment.² Studies have been conducted to predict disease onset for MCI patients based on brain structure volume features such as temporal lobe atrophy,³ corpus callosum atrophy,⁴ and hippocampus subfield volume.⁵ In addition to brain structure volume features, other MRI-based imaging features have also been studied as early biomarkers of AD, such as hippocampal texture⁶ and gray/white matter contrast difference.⁷ Coupé et al.⁸ introduced an MRI-based hippocampal grading score based on patch-based image similarity comparison between test images and training templates to predict AD onset seven years before the conversion. Furthermore, an increasing number of recent studies has achieved improvement in compute-aided AD diagnosis through the extraction of multi-dimensional imaging-based features using statistical or machine-learning approaches.⁹⁻¹³ However, most of the previous studies in the literature only use single timepoint images to predict the onset of AD. However, the pathological progression of AD is a longitudinal process, which has been analogous to an accelerated aging process with structural change.¹⁴ Therefore, temporal features, which are missing in the single timepoint analysis, consist of a critical source of information to be taken into account.

Studies have attempted to take advantage of the information embedded in the longitudinal pathological progression as revealed in the images, e.g., the volume changes over time of AD-specific structures, such as the hippocampal atrophy, and the ventricle expansion. Zhang and Shen¹⁵ proposed to use multi-modal image (MRI + PET) and longitudinal data from multiple timepoints using a sparse linear model for feature selection. Several studies have been proposed to capture the temporal changes through longitudinal image registration. Fiot et al.¹⁶ introduced a classifier for MCI converters versus non-converters based on the logistic loss and spatial regularization of the longitudinal hippocampal deformation between two timepoints. Later on, Lorenzi et al.¹⁷ introduced a longitudinal model of disease-specific morphological age based on the longitudinal deformation from baseline to follow-up scans. More recently, Sun et al.¹⁸ achieved high accuracy in classifying MCI converters against non-converters for up to three years by leveraging the stationary velocity field from the non-rigid registration between the baseline and follow-up images.

Some other studies have proposed to model the longitudinal disease trajectories using linear mixed-effect model,¹⁹⁻²² in which case the individualized variations are considered as random effects either toward the intercept or the slope of the longitudinal scan time in the linear model. However, the nature of the linear model in these studies relies on the underlying assumption that the structural change is linear.²² On the other hand, brain structural atrophy rate is age-dependent, and thus a linear model would be sub-optimal.²³

Current existing longitudinal analysis methods using conventional linear least-squared model or linear mixed model are unable to completely capture the complexity of longitudinal disease progression trajectories, especially under the presence of complicated non-linear trend. Moreover, it is difficult to estimate the change in the longitudinal trajectories when the data are sparsely and irregularly observed or even missing.²⁴ To characterize the variations in the longitudinal trajectories and to capture its relationship with the MCI-to-AD conversion, we propose to use functional principal component analysis (FPCA) and functional linear regression (FLR) to fit the trajectories and produce prediction using machine learning techniques.

FPCA is a data-driven methodology for detecting the principal direction of variations in longitudinal trajectories and thereby converting the infinite-dimensional curves into low-dimensional vectors. Depending on how the data points are sampled from the curves, various methods for conducting FPCA have been proposed.²⁵⁻²⁷ Of particular interest is the case where the data are sparsely sampled, as it relates directly to our application. For such sparse and irregular

longitudinal data, Yao et al.²⁸ proposed the principal analysis by conditional expectation, and Hall et al.²⁹ established its mathematical validity in terms of asymptotical statistical properties.

When a random process is sampled at multiple timepoints, the data can be viewed as a function of time and are generally referred to as the functional data. FLR is a novel regression framework for handling functional predictors. It can be used to model the relationship between the response and the functional covariates and to predict a scalar response from the observed data. Theoretical properties of FLR have been well established in the literature. Yao et al.³⁰ studied FLR for sparse data and proposed a flexible method for prediction based on FPCA. Gertheiss et al.³¹ developed a new type of functional regression model by extending the classical functional regression to accommodate variation in subject-specific trends. A systematic review of FLR can be found in Morris.³²

In this paper, the FPCA is used for extracting the main sources of variations of the longitudinal disease progression trajectories, and the FLR is used for predicting the AD. The FPCA allows us to summarize the information in the nonlinear longitudinal trajectories into a few simplistic FPC scores. These FPC scores enable us to build functional regression models and can be used as pathological-progression related features to apply machine learning methods for predicting the AD. We evaluated the application by measuring the accuracy of predicting the disease onset one and two years in advance.

2 Methods

2.1 Experimental Data and Data Preparation

A total of 7656 T1-weighted MPRAGE brain MRI images were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database,^{33,34,34-37} which contains longitudinally scanned images from 1727 subjects. The subjects' demographic information, along with the diagnoses at all the scanning timepoints were provided in the ADNI database.³⁷ Each patient was diagnosed as one of the three categories: CN, MCI, or AD. Each subject's brain structural volumes were measured around every six months in the first two years and subsequently every 12 months until the end of their follow-up periods.

2.1.1 FreeSurfer structure segmentation and volume extraction

The brain MRI was segmented into 87 laterally separated anatomical structures using FreeSurfer processing framework (version 5.3)³⁸ following the pipeline of: (1) affine registration to the MNI205 probability atlas; (2) B1 bias field correction; (3) non-rigid registration to correct for the local morphological variance; (4) expectation-maximization-based structural labeling and parcellation.^{39,40} A total of five AD-related regions of interest (ROI), i.e., hippocampus, lateral ventricle, amygdala, entorhinal, and parahippocampal, were included in the study.

Under the functional regression model framework, longitudinal trajectories of biomarker measurements within the first 24 months are used for constructing the functional predictors. Binary outcomes of whether the diagnosed status of the individual converts from MCI to AD as at month 36 (one-year-ahead prediction) and month 48 (two-year-ahead prediction) are used as the dependent variables in the functional regression model, respectively. It is worth noting that the actual time of follow-up and diagnosis is not exactly 24/36/48 months and may differ by the nominal time by a few days. For more accurate modeling, the actual time of diagnosis rather than the nominal one is used in our analysis.

2.1.2 Data harmonization

The general linear model (GLM) is used to correct for the residual effects of field strength, sex, and total intracranial volume (TIV).⁴¹ We use the field strength, sex, and TIV as predictors, and fit the GLM to the structural volume data on the reference group consisting of subjects diagnosed with CN only,^{8,41} and calculate the standardized residual between the predicted and actual volumes (also known as W-score, defined as $W_i = (\varepsilon_i - \mu_{\varepsilon_{CN}}) / \sigma_{\varepsilon_{CN}}$).⁴²⁻⁴⁴ The demographic

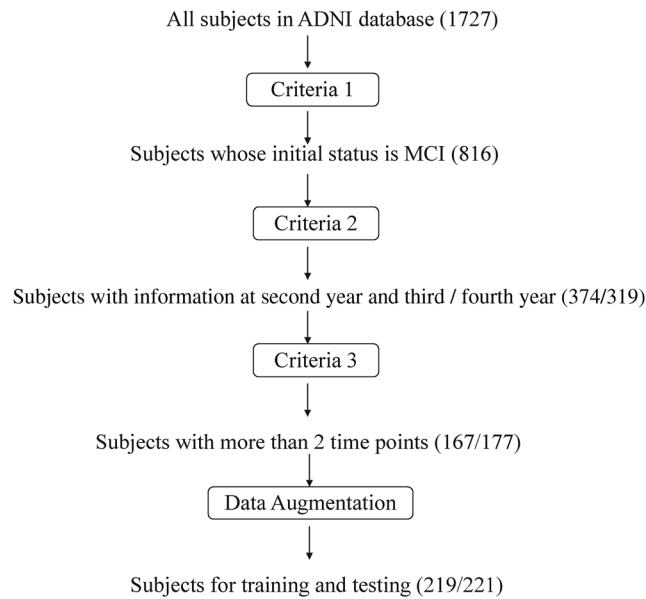


Fig. 1 Flowchart of the data inclusion for one-year-ahead and two-year-ahead analysis. The number in the parenthesis is the number of subjects of the one-year-ahead and two-year-ahead studies.

information (i.e., age and sex)³⁷ and the scanner-specific information (i.e., field strength)³⁴ are obtained from the ADNI database alongside with the T1 image data. The TIV is estimated by the multi-atlas label-fusion method.⁴¹ Volumes of the five brain ROI are subsequently corrected for the joint effects related to age, sex, and TIV using the respective regression models. We refer to the resulting corrected values as “GLM-corrected volumes”.

2.1.3 Training data augmentation

The final study subjects are selected based on the following criteria: (1) The initial status of the subjects should be MCI. (2) Subjects should have information at the end of the second and the third year (for one-year-ahead analysis) or the fourth year (for two-year-ahead analysis). (3) Subjects should have more than two-time points. Figure 1 shows the flowchart of the data inclusion.

ADNI data include longitudinal scans for each individual subject. At the time when this study was conducted, the maximum period between the baseline and follow-up scans was 10 years. We augment the training data using a sliding window of 12 months along an individual’s temporal trajectory. For example, consider an individual whose baseline and follow-up scans are collected on 0, 6, 12, 18, 24, 36, and 48 months, and the one-year-ahead prediction (i.e., month 36) given the first 24 months’ trajectory (i.e., month 0, 6, 12, 18, and 24) is of interest. We may augment the data by treating month 12 as the starting time point and moving the 12-month sliding window ahead by 12 months, thereby creating an augmented record that is based on the trajectory of month 12, 18, 24, 36, and one-year-ahead prediction on month 48. Table 1 shows the detailed breakdown of the individuals classified as non-converter (MCI → MCI) and converter (MCI → AD) as at month 24 for one-year-ahead and two-year-ahead predictions.

2.2 Longitudinal Disease Trajectory Feature Extraction

Functional data analysis is a useful tool for analyzing sparse and noisy longitudinal data. In particular, the FPCA is used for extracting the primary mode of variations in the longitudinal trajectories, thereby decomposing the underlying stochastic process into several orthogonal functional components. These uncorrelated components, referred to as the FPC, can be further used as a foundation for functional regression modeling.

Table 1 Breakdown of number of subjects in the original and augmented data sets for one-year-ahead and two-year-ahead predictions.

Diagnosed status	Number of subjects			
	One-year-ahead		Two-year-ahead	
	Original	Augmented	Original	Augmented
Non-converter (MCI → MCI)	146	184	142	174 [t]
Converter (MCI → AD)	21	35	35	47
Total (converter + non-converter)	167	219	177	221

2.2.1 Functional principal component analysis

We propose to analyze the trajectories of the GLM-corrected volumes within the paradigm of functional data analysis. The first step is to extract the FPCs from each group of volume trajectories. We model the trajectories of a certain brain region as independent realizations from a stochastic process $X(t)$ and let $X_i(t)$ denote the trajectory realization of the i th subject. Let $\mu(t) = E(X(t))$ and $G(s, t) = \text{Cov}(X(s) - \mu(s), X(t) - \mu(t))$ denote the mean function and the covariance function, respectively. Based on the Karhunen–Lovève decomposition, $X_i(t)$ can be expressed as

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t), \tag{1}$$

where $\phi_k(t)$ is the k 'th eigenfunction, and ξ_{ik} is the associated FPC score for the i 'th subject. The eigenfunctions should satisfy

$$\int_{\mathcal{T}} \phi_k(t) \phi_j(t) dt = \delta_{kj},$$

where $\delta_{kj} = 1$ if $k = j$ and 0 otherwise.

The FPC score is defined as

$$\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t)) \phi_k(t) dt. \tag{2}$$

The magnitude of ξ_{ik} represents the degree of similarity between the $X_i(t) - \mu(t)$ and the eigenfunction $\phi_k(t)$. The mean and variance of the distribution of ξ_{ik} are $E(\xi_{ik}) = 0$ and $\text{Var}(\xi_{ik}) = \lambda_k$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$.

By the Mercer's theorem, the covariance function $G(s, t)$ can be expressed as

$$G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t). \tag{3}$$

As it would be unrealistic to estimate an infinite number of eigenfunctions, in reality, $X_i(t)$ is usually well approximated by retaining only the first K leading eigenfunctions and the related FPC scores,

$$X_i(t) = \mu(t) + \sum_{k=1}^K \xi_{ik} \phi_k(t). \tag{4}$$

In the case of sparse data with possible measurement error, the eigenfunctions and FPC scores can be obtained by applying the principal components analysis through the conditional expectation (PACE) method.²⁸ The eigenfunctions and FPC scores under the PACE method are

computed as follows. The mean function $\mu(t)$ is first estimated by pooling the data points from all the trajectories together and applying local linear smoothing; the covariance function $G(s, t)$ is estimated by a two-dimensional kernel smoother. Let $\hat{\mu}(t)$ and $\hat{G}(s, t)$ denote the estimated mean trajectory and the estimated smoothed covariance surface. The eigenfunction $\phi_k(t)$ satisfies the eigenequation

$$\int_{\mathcal{T}} \hat{G}(s, t)\phi_k(s)ds = \lambda_k\phi_k(t), \tag{5}$$

with the constraints $\|\phi_k\|^2 = 1$ and $\langle \phi_k, \phi_j \rangle = 1$ if $k = j$, and 0 otherwise. The solution to such an eigenequation $\hat{\phi}_k(t)$ can be found by applying spectral decomposition on the discretized covariance function $\hat{G}(s, t)$.

Let $Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$ be the measurement at the j 'th time point ($j = 1, \dots, n_i$) of the i 'th subject's curve, and ϵ_{ij} is assumed to be a random error term following a normal distribution with mean 0 and variance σ^2 . Let $\hat{\phi}_{ik}$ and $\hat{\mu}_i$ denote the vectors of values of $\hat{\phi}_k(t)$ and $\hat{\mu}(t)$ evaluated at time points t_{ij} , and let $\hat{\mathbf{G}}_i$ denote the matrix of values of $\hat{G}(s, t)$ evaluated at the two-dimensional grid consisting of t_{ij} .

The FPC score of the i 'th subject and the k 'th eigenfunction is computed from the conditional expectation

$$\hat{\xi}_{ik} = \hat{\mathbf{E}}(\xi_{ik}|\mathbf{Y}_i) = \lambda_k \hat{\phi}_{ik}^T \hat{\Sigma}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \hat{\mu}_i),$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$, $\hat{\Sigma}_{\mathbf{Y}_i} = \hat{\mathbf{G}}_i + \hat{\sigma}^2 \mathbf{I}_{n_i}$.

2.2.2 Feature selection and regularized-model-based analysis

The classic functional regression model aims at discovering the relationship between the functional data, i.e., the longitudinal trajectories, and a scalar response, i.e., the binary status of AD in our context. As the eigenfunctions are orthogonal to each other, the resultant product of the regression coefficient function and the longitudinal trajectory can be transformed and simplified as a linear combination of the FPC scores. Therefore, the FPC scores can be directly used as predictors in the usual framework of the generalized linear model.

Let D_i denote the binary AD status of the i 'th subject, and let \mathbf{Z}_i denote the vector of scalar predictors, for example, \mathbf{Z}_i may consist of a subject's age and sex in our context. The functional regression model is constructed as

$$\text{logit}\{\Pr(D_i = 1|X_i(t), \mathbf{Z}_i)\} = \int_{\mathcal{T}} X_i(t)\beta(t)dt + \mathbf{Z}_i^T \boldsymbol{\gamma}, \tag{6}$$

where $\beta(t)$ is the function of the regression coefficient for the longitudinal trajectories and $\boldsymbol{\gamma}$ is the vector of coefficients for the scalar predictors.

By expanding on the basis of eigenfunctions $\{\phi_k(t), k = 1, \dots, K\}$, the regression coefficient function $\beta(t)$ can be expressed as

$$\beta(t) = \sum_{k=1}^K c_k \phi_k(t),$$

where $c_k = \int_{\mathcal{T}} \beta(t)\phi_k(t)dt$ is the basis coefficient and can be subsequently estimated from the generalized linear model. Therefore, we may rewrite Eq. (6) as

$$\text{logit}\{\Pr(D_i = 1|X_i(t), \mathbf{Z}_i)\} = \sum_{k=1}^K c_k \xi_{ik} + \mathbf{Z}_i^T \boldsymbol{\gamma}.$$

We first estimate the FPC scores ξ_{ik} using the PACE method, and use the estimated FPC scores as the predictors in the generalized linear model.

Using the FPC scores as predictors, we are able to recover the relationship between the MCI/AD conversion and the functional shapes of the longitudinal volumes change trajectories. Not all the predictors would be of statistical significance or predictive value. As a variable selection procedure, we apply the LASSO algorithm to identify the influential variables and thereby predict the MCI/AD conversion status in a more accurate fashion.

2.3 Longitudinal Disease Prediction and Cross-Validation

After calculating the FPC scores for each trajectory, we adopt a stratified five-fold cross-validation approach to assess the prognostic accuracy of the functional regression model for classifying the patient into potential converter (MCI \rightarrow AD) and non-converter (MCI \rightarrow MCI). Specifically, in the proposed cross-validation approach, data are randomly partitioned into five folds and individuals are stratified, i.e., each partition is equally sized, and the ratio of the converter/non-converter is roughly the same across all the folds. Within the five-fold cross-validation framework, the number of subjects in a training data set is roughly 175 and 177 for one-year-ahead and two-year-ahead prediction task, respectively. The number of subjects in the testing data set is 44. The analysis is repeated five times to obtain stable estimates of the prediction performance metrics. Among these five partitions, one partition is kept to be used separately for testing the model, and the rest of the partitions are used for training the model. The cross-validation process is repeated five times, with each partition used as the testing set. Because multiple augmented data record may originate from the same individual, data contamination may occur if these records appear both in the training and testing data set, leading to false-positive improvement in the prediction performance. To ensure the objectiveness of the cross-validation process, we avoid potential data contamination by enforcing the rule that augmented data points can only appear in the training set and an individual can appear either in the training set only or in the testing set only, but not both. We implement such a rule by partitioning the original subjects of the augmented data records into five groups and assigning the corresponding records to each cross-validation fold. The classification performance is assessed in terms of the area under the receiver operating characteristic curve (AUC).

FPC scores can be calculated from a reference group that consists of either the whole or a subset of the population. To better understand the effect of the choice of reference group on the accuracy of disease prediction, we perform six parallel experiments with different combinations of diagnosis groups to calculate the FPC scores for the training data in each fold of the cross-validation. The choice of the reference group in the six experiments are: (1) CN only, (2) AD only, (3) MCI only, (4) CN + MCI, (5) AD + MCI, and (6) CN + MCI + AD. For each reference group, we fit an FPCA model to extract the first five FPCs of the volume trajectories of the five ROIs. The resultant FPCA models are used for computing the FPC scores of the subjects in the MCI group, which are used as the covariates in the functional regression model with LASSO.

Furthermore, we also compare the prediction accuracy of the same classifier with two types of input features: (1) the proposed longitudinal FPC score and (2) the single-timepoint structure volume. A common practice in single-timepoint-based prediction is to train the classifier using only the CN and AD group and then test on the MCI group so that the classifier learns from the two extreme sides of the disease spectrum (CN and AD) and then test on the more difficult cases in the MCI group.^{3,6,8,45} For the purpose of a side-to-side comparison, we also choose the CN and AD groups for training with the longitudinal FPC score feature.

3 Results

3.1 Longitudinal Structural Change along with Disease Progression

We first investigate the overall ROI volume change along time, both in terms of structural atrophy and ventricle expansion. We plot the mean trajectories of the converter and non-converter groups in Fig. 2. The average trajectories show both Atrophy (A-D) and ventricle expansion (E), although considerable variation exists in the longitudinal trajectories of the five brain regions, as shown in Fig. 2. Interestingly, comparison between the mean longitudinal trend of the non-

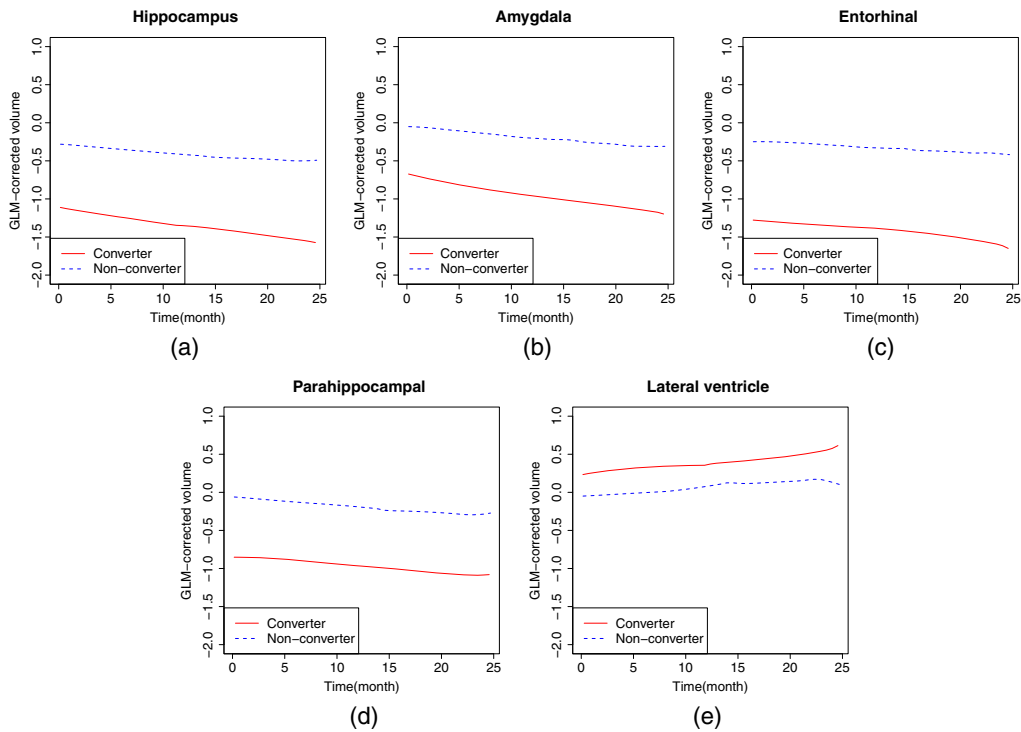


Fig. 2 Mean curves of the GLM-corrected volumes of the five brain regions for AD (solid) and non-AD (dashed) subjects.

converter group (blue line, Fig. 2) with that of the converter group (red line, Fig. 2) reveals a trend of not only smaller atrophy but also at a relatively reduced atrophy rate in hippocampus, amygdala, entorhinal, and parahippocampus (A-D). Similarly, compared with the converter group, the ventricle expansion (E) is smaller and progresses with a reduced rate for the non-converter group.

3.2 Longitudinal FPC Feature Extraction for Early AD Prediction

We apply the FPCA to the longitudinal trajectories of the GLM-corrected ROI volumes to further explore the in-depth information for the volume changes over time that are embedded in the disease progression trend. The PACE method is used to extract the functional principal components from the pool of longitudinal measurements in each group.³⁰

Figure 3 exhibits the plots of the trajectories of the first three eigenfunctions (or FPCs) for the parahippocampal. A particular aspect of interest of the FPC curve is its sign, which indicates the direction of variation in the change of the trajectory along time. To be specific, the first FPC is constant above zero. It represents that the main source of variation among the parahippocampal comes from the weighted average of their volume trajectories. The second FPC crosses the x axis once: it is positive in $[0, 15.5]$ and negative afterward, representing the change in the volume trajectory after the 15.5 month. Moreover, the magnitude of change after the 15.5 month appears to be larger, indicating a sharper shift in the volume curve. The third FPC crosses the x axis twice and is primarily positive in $[2.9, 12.1]$ and negative in the other intervals, which can be interpreted as the difference in the volume trajectory during $[2.9, 12.1]$ and the other time intervals. It should be noted that the estimation of the FPC near the boundary (around 24 months) could be subject to larger variances due to lesser data points on the time grid, and thus we are primarily focused on changes in the curve inside the boundary region. The FPCs of the other four ROIs show similar longitudinal patterns.

Table 2 shows the proportions of variation among the population that are explained by the top five FPCs. For example, the first five FPCs of hippocampus volume trajectories explain 95.56%, 3.40%, 0.70%, 0.13%, and 0.11% of the total variation, respectively. The first FPC

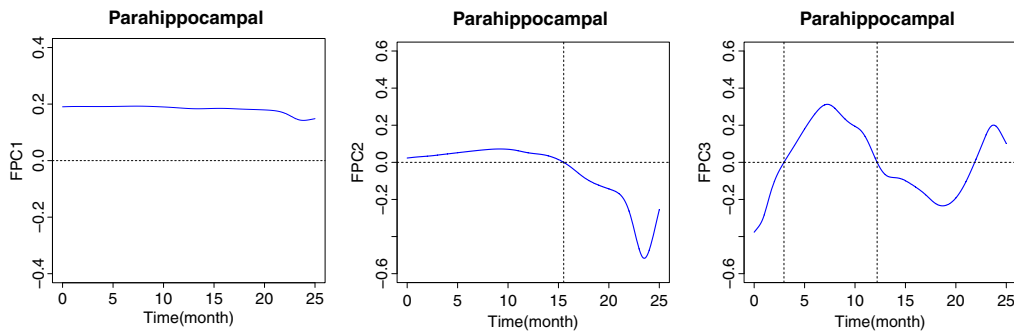


Fig. 3 Breakdown of the first three FPCs (from left to right) of longitudinal trajectories of the GLM-corrected volumes of the parahippocampal.

Table 2 Proportion of variation (in percentages) explained by the top five FPCs.

Predictors	FPC1	FPC2	FPC3	FPC4	FPC5
Hippocampus	95.56	3.40	0.70	0.13	0.11
Lateral ventricle	95.82	2.96	0.71	0.28	0.12
Amygdala	95.70	3.85	0.22	0.12	0.04
Entorhinal	96.86	2.31	0.65	0.08	0.04
Parahippocampal	95.74	3.83	0.21	0.10	0.06

captures more than 95% of the total variation of all the evaluated structures, with each following FPC capturing the largest proportion of the remaining total variation.

3.3 Analysis of Feature Importance and Feature Selection

We utilize the extracted FPC scores of the longitudinal structural volume change as the features for predicting whether an MCI subject would develop AD (converter) at a later time point (either one year later or two years later). For each subject, a total of $5 \times 5 = 25$ FPC scores (five structures for each subject, and the first five FPCs for each structure) are used as the covariates for the one-year-ahead and two-year-ahead AD prediction. As the number of predictors is relatively large, conventional GLM, e.g., logistic regression, would either fail to converge or provide estimates with high variances. Feature selection is imperative and we resort to the LASSO algorithm to find the FPC scores that have the most predictive value on distinguishing between the converter and the non-converter.

In the LASSO model, the identified influential FPC scores for the converter/non-converter classification are grouped according to the brain region they belong to and are shown in Fig. 4. The variable importance scores of FPCs from the same category are grouped and summed for comparison, and the scores are normalized to the range of 0 and 1. The FPC scores of the parahippocampal are identified as the most influential features, and their variable importances are significantly higher than those of the remaining regions.

3.4 Model Evaluation and Cross-Validation for AD Classification

In this section, results of model performance in terms of the predictive power of classifying the MCI-AD converter and MCI-MCI non-converter are presented. The AUC averaged across the five-folds cross-validation is used as the summary measure for the overall predictive power.

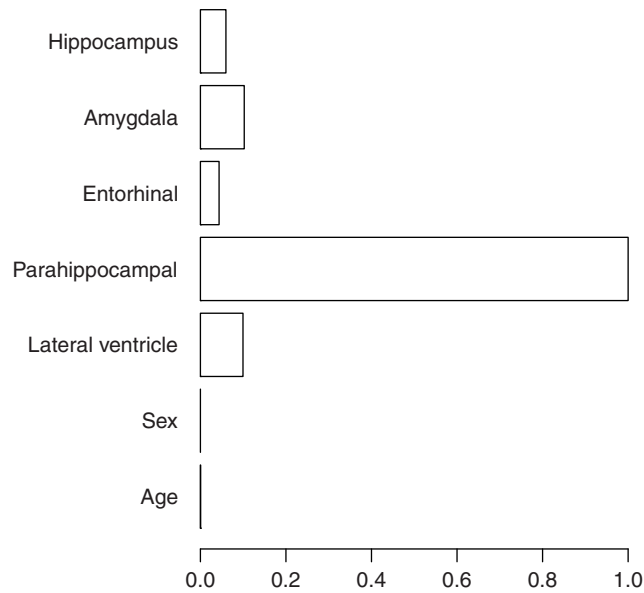


Fig. 4 Variable importance of the FPC scores of the five brain regions and the scalar predictors (age and sex) under the LASSO model.

Table 3 Average classification performance on the testing set of MCI group for models using different types of feature as predictor (first column) and different groups as source of feature (second column). Paired *t*-test was conducted between each longitudinal FPC-based experiment to the baseline volume feature experiment. Multiple comparison was corrected with false discovery rate (FDR) set to 0.05.

Feature type	Feature source	AUC (SD)	
		12 months	24 months
Baseline volume	CN + AD	0.732(0.140)	0.717 (0.157)
Longitudinal FPC	CN + AD	0.802(0.145)***	0.816 (0.094)
Longitudinal FPC	CN	0.726(0.173)	0.789 (0.110)
Longitudinal FPC	AD	0.819(0.127)***	0.800 (0.100)
Longitudinal FPC	MCI	0.740(0.059)	0.775 (0.108)
Longitudinal FPC	CN + MCI	0.790(0.128)**	0.774 (0.138)**
Longitudinal FPC	MCI + AD	0.818(0.127)***	0.802 (0.090)
Longitudinal FPC	CN + MCI + AD	0.814(0.135)***	0.811 (0.091)

*FDR-corrected *P*-value <0.05
 **FDR-corrected *P*-value <0.01
 ***FDR-corrected *P*-value <0.001.

Table 3 shows the average AUC in the testing set for different models and predictors. We first compare the approaches using longitudinal FPCA against the one based on the baseline volume. The training data consist of the stable AD versus the stable CN subjects, which has been shown to improve the classification power compared with training using MCI subject only.^{3,6,8,45} Table 3 shows further improvement in classification accuracy when incorporating different reference groups to calculate the FPC scores. Three reference groups are considered: stable AD, stable CN, and MCI (converter and non-converter). To have a side-to-side comparison, we used the same folds of training/testing split across all the models using longitudinal FPC scores as

features. When using the MCI group as the training set to derive the longitudinal FPC features, we include the longitudinal volume information from multiple timepoints and improve the predictive accuracy as indicated by an increase in the AUC. It is evident that the classifier trained with the FPC-based longitudinal features outperforms the single-timepoint volume features. Using the AD group to train the FPC model results in a significant improvement in the classification power compared with the model with CN- or MCI-based FPC features. Inclusion of the CN and MCI reference groups on top of the AD data leads to mild improvement. Moreover, it is observed that there is no evident difference between the predictive accuracies of the one-year-ahead and two-year-ahead predictions.

4 Discussion and Conclusion

Dementia of the Alzheimer type is a prevalent disease worldwide with no definitive cure available yet, and it is therefore important to detect and predict the disease onset earlier to create enough time-windows, which are precious for disease intervention. Given the fact that the risk of disease onset increases with age, it would be beneficial to grasp subtle time-dependent changes in the brain structures within a short period of follow-up in the early stage of the disease before the symptoms manifest.

In this paper, we present a novel framework using the FPCA to extract features from MRI-derived information collected from multiple timepoints. Our study shows promising results and demonstrates improved prediction power using FPC-derived longitudinal features compared to naive features based on single timepoint.

4.1 Importance of Incorporating Longitudinal Information for Early Disease Prediction

Longitudinal information is important for early prediction of potential future disease onset. However, most of the current literature in AD study only use the single timepoint cross-sectional data to build the classifier. It has been shown that incorporating longitudinal information can improve the classification power for determining disease progression or for identifying potential treatment effect.⁴⁶ In this study, we showcase the application of extracting the longitudinal disease progress information (in terms of MCI-to-AD conversion) through the FPCA, and demonstrate improved classification performance over using the single-timepoint volume feature.

4.2 Advantage of FPCA over other Methods

Aimed at incorporating the trajectory information of the volume into the prediction of AD, we suggest a regularized function regression approach based on FPCA. The comparison results indicate that the proposed trajectory-based method leads to an increase in the predictive accuracy over the conventional single timepoint based models. The reason for such an improvement is as follows. As Eq. (6) suggests, the integral of the product between the trajectory value and the coefficient function is a versatile and comprehensive summary of the cumulative effect of the volume trajectory in distinguishing between the converter versus non-converter. The FPCA is able to extract the information from the longitudinal trend and has shown promising successful applications in many fields, including biomedical and clinical studies,^{47,48} ecology,⁴⁹ psychology,⁵⁰ and econometrics.⁵¹

4.3 Choice of the Reference Group for FPC Score

In its essence, FPCA is a dimension reduction approach to project the longitudinal information of the data into a more representative low-dimensional coordinate system. Therefore, the choice of the reference group for constructing such a coordinate system, i.e., reference sample data, is crucial. In the results of our study, as shown in Table 3, choosing the MCI-only group as the reference group for FPCA results in the least discriminative features sets, compared to the CN-only and AD-only reference group. This can be explained by the fact that MCI group consists of

subjects who are in between the healthy control and diseased population. Therefore, their coordinates in the projected FPC space would have less predictive variation than the ones in the projected FPC space based on CN and AD group's longitudinal trajectories. As a result, the coordinates in the projected FPC space based on the MCI group would be less discriminative, resulting in a sub-optimal classification result.

In comparison, choosing AD group as the reference sample to conduct the FPCA significantly improves the classification power for both the 12 months and 24 months predictive tasks, but interestingly, choosing the CN group as the reference sample shows similar classification results compared to the case where MCI-group is used as the reference sample. This observation may indicate more similarity of the MCI group to the CN group when comparing their longitudinal disease progression trend using FPCA. To sum up, the above observations indicate that the feature of longitudinal volume change demonstrated larger separation between the MCI group and the AD group compared to the features based on absolute structure volume. Such insights demonstrate the advantage of longitudinal-based analysis to achieve the task of distinguishing MCI subjects with AD patients. Furthermore, it would be interesting to further divide the MCI group into converter and non-converter, and to include the converter-MCI subjects as a separate reference group, as suggested by Popuri et al.⁵² It is possible to investigate whether such a stratified group definition would better reflect the more detailed difference in the longitudinal information within each diagnosis subgroup.

4.4 *Difference between the Proportion of Total Variance and the Importance of Variable for Classification*

Table 2 shows the importance of FPC in terms of the proportion of the population's variation for each structural ROI biomarker that the corresponding FPC can capture. On the other hand, Fig. 4 shows the importance of the features that can classify and explain the difference between the converter and the non-converter groups. It is worth noting that, although the first FPC (FPC1) always explains the largest proportion of populational variation for each feature (as shown in Table 2), they are not necessarily the most important features responsible for capturing the group difference for classification. Such a difference is indicative of the fact that the populational variation on that particular FPC may contain more intra-group variation than the inter-group variation.

4.5 *Other Methods of Early Disease Prediction*

Several previous studies have attempted to extract longitudinal disease trajectory for AD prediction. The temporal-image-registration between two or more time-points¹⁶⁻¹⁸ are prone to registration error, require heavy computation, and are usually restrained the analysis within a single structure. The attempt to construct longitudinal features by simply combining all the coefficient at all the timepoints¹⁵ suffers from overfitting due to an overly large number of parameters. On the other hand, linear disease trajectory models¹⁹⁻²³ assume that the longitudinal structural change are linear, resulting in model under-fitting.

Comparatively, our method defines a fixed number of longitudinal features (= 5 FPC scores per ROI), which captures the most significant five principal functional components of the longitudinal trajectories. Furthermore, comparing with our longitudinal feature selection method, Zhang and Shen¹⁵ used a feature weight vector with a group regularization constraint, and performed the linear feature selection based on the feature weights.

Recent studies have also explored the utility of deep neural networks to incorporate longitudinal information for improving early disease prediction of AD. Bhagwat et al.⁵³ proposed longitudinal siamese neural network with novel architectural modules for combining multimodal data from two timepoints to predict clinical scores. Recently, Cui and Liu⁵⁴ adopted a recursive neural network to incorporate longitudinal information with cascaded three bi-directional gated recurrent unit, and combined it with the convolutional neural networks to improve the longitudinal disease prediction. It would be interesting to compare the statistically modeled longitudinal method and the deep-learning-based approach, and seek ways to combine the two.

4.6 Limitation of the Current Study

In this study, we investigate the improvement of FPCA by incorporating the longitudinal data from the first 24 months with a total of four timepoints (six months apart in the first three timepoints, and 12 month apart between the third and fourth timepoint) and demonstrate significant improvement in terms of disease prediction compared with the classification method that utilizes only the single-timepoint data. A natural extension of the current approach would be to investigate the effect of the number of follow-up timepoints incorporated on the classification performance of the FPCA model. The effect of different number of the follow-up timepoints from longitudinal data can be accessed in a step-wise fashion by gradually increasing the number of follow-up timepoints.

Furthermore, the FPCA is a statistical method that can capture the overall sample variations of the longitudinal trends in disease progression along multiple timepoints within the study period. In this study, our approach is to recover the FPC using a limited number of early timepoints to predict the probability of future disease onset through classification. Taking a step further, an alternative approach would be to directly predict the future measurement of the features, the structural volume in our case, as a surrogate representation of the longitudinal trend of disease progression. The FPCA in this case essentially serves as a statistically sound method for missing-data-recovery. In this setup, data across all the timepoints can be included, leading to the construction of a more comprehensive statistical model that takes a full representation of the whole data set.

Finally, the histopathological characterization of AD defined as the presence of the brain pathologies such as the beta-amyloid ($A\beta$) plaques, neurofibrillary tangles, and neuritic plaques. On the other hand, clinical AD is restricted to demented subject only, therefore doesn't include those patients with AD-related brain pathology but without clinical syndrome, the latter of which has been proposed to be categorized as preclinical AD or asymptomatic AD. In this study, we mainly focused on the clinically diagnosed AD, rather than the cases with asymptomatic AD.

Disclosures

There is no conflict of interest to declare from all authors.

Acknowledgments

Funding for this research is gratefully acknowledged from National Science Engineering Research Council (NSERC), Canadian Institutes of Health Research (CIHR), Brain Canada Foundation, Pacific Alzheimer's Research Foundation, the Michael Smith Foundation for Health Research (MSFHR), Alzheimer Society Research Program from Alzheimer Society of Canada and Alzheimer Society of British Columbia, and the Canadian Institute on Aging. We thank Compute Canada for the computational infrastructure provided for the data processing in this study. Data collection and sharing for this project is funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda 31Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada.

References

1. K. A. Johnson et al., "Brain imaging in Alzheimer disease," *Cold Spring Harb. Perspect. Med.* **2**(4), a006213 (2012).
2. J. Ye et al., "Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data," *BMC Neurol.* **12**, 46 (2012).
3. A. Chincarini et al., "Automatic temporal lobe atrophy assessment in prodromal AD: data from the DESCRIPA study," *Alzheimer's Dementia* **10**(4), 456–467 (2014).
4. S. Elahi et al., "Corpus callosum atrophy rate in mild cognitive impairment and prodromal Alzheimer's disease," *J. Alzheimer's Disease* **45**(3), 921–931 (2015).
5. W. Khan et al., "Automated hippocampal subfield measures as predictors of conversion from mild cognitive impairment to Alzheimer's disease in two independent cohorts," *Brain Topogr.* **28**, 746–759 (2015).
6. L. Sorensen et al., "Early detection of Alzheimer's disease using MRI hippocampal texture," *Hum. Brain Mapp.* **37**(3), 1148–1161 (2016).
7. A. L. Jefferson et al., "Gray and white matter tissue contrast differentiates mild cognitive impairment converters from non-converters," *Brain Imaging Behav.* **9**, 141–148 (2015).
8. P. Coupé et al., "Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: toward an early individual prognosis," *Hum. Brain Mapp.* **36**(12), 4758–4770 (2015).
9. V. Karami, G. Nittari, and F. Amenta, "Neuroimaging computer-aided diagnosis systems for Alzheimer's disease," *Int. J. Imaging Syst. Technol.* **29**(1), 83–94 (2019).
10. K. Popuri et al., "Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases," *Hum. Brain Mapp.* **41**, 4127–4147 (2020).
11. D. Ma et al., "Differential diagnosis of frontotemporal dementia, Alzheimer's disease, and normal aging using a multi-scale multi-type feature generative adversarial deep neural network on structural magnetic resonance images," *Front. Neurosci.* **14**, 853 (2020).
12. E. Yee et al., "Construction of MRI-based Alzheimer's disease score based on efficient 3D convolutional neural network—comprehensive validation on 7,209 multi-centre dataset," *J. Alzheimer's Disease* **79**(1), 47–58 (2021).
13. D. Ma et al., "Blinded Clinical Evaluation for Dementia of Alzheimer's Type Classification Using FDG-PET: A Comparison Between Feature-Engineered and Non-Feature-Engineered Machine Learning Methods," *J. Alzheimers Dis.* **80**(2), 715–726 (2021).
14. C. Davatzikos et al., "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiol. Aging* **32**(12), 2322.e19–2322.e27 (2011).
15. D. Zhang and D. Shen, "Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers," *PLoS One* **7**, e33182 (2012).
16. J. B. Fiot et al., "Longitudinal deformation models, spatial regularizations and learning strategies to quantify Alzheimer's disease progression," *Neuroimage Clin.* **4**, 718–729 (2014).
17. M. Lorenzi et al., "Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images," *Neurobiol. Aging* **36**(Suppl. 1), S42–S52 (2015).
18. Z. Sun et al., "Detection of conversion from mild cognitive impairment to Alzheimer's disease using longitudinal brain MRI," *Front. Neuroinf.* **11**, 16 (2017).
19. P. Schulam and S. Saria, "A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure," in *Adv. Neural Inf. Process. Syst.*, pp. 748–756 (2015).
20. J. Futoma et al., "Predicting disease progression with a model for multivariate longitudinal clinical data," in *Mach. Learn. Healthcare Conf.*, pp. 42–54 (2016).
21. D. G. Roche, V. Careau, and S. A. Binning, "Demystifying animal 'personality' (or not): why individual variation matters to experimental biologists," *J. Exp. Biol.* **219**, 3832–3843 (2016).
22. H. Lee et al., "Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements," *NeuroImage* **184**, 555–565 (2019).
23. X. Hua et al., "Sex and age differences in atrophic rates: an ADNI study with $n = 1368$ MRI scans," *Neurobiol. Aging* **31**(8), 1463–1480 (2010).

24. Y. Nie, Y. Yang, and J. Cao, "Recovering the underlying trajectory from sparse and irregular longitudinal data," arXiv:1809.06462 (2018).
25. J. Dauxois, A. Pousse, and Y. Romain, "Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference," *J. Multivariate Anal.* **12**(1), 136–154 (1982).
26. J.-T. Zhang and J. Chen, "Statistical inferences for functional data," *Ann. Stat.* **35**, 1052–1079 (2007).
27. M. Benko, W. Härdle, and A. Kneip, "Common functional principal components," *Ann. Stat.* **37**, 1–34 (2009).
28. F. Yao, H.-G. Müller, and J.-L. Wang, "Functional data analysis for sparse longitudinal data," *J. Am. Stat. Assoc.* **100**(470), 577–590 (2005).
29. P. Hall, H.-G. Müller, and J.-L. Wang, "Properties of principal component methods for functional and longitudinal data analysis," *Ann. Stat.* **34**, 1493–1517 (2006).
30. F. Yao, H.-G. Müller, and J.-L. Wang, "Functional linear regression analysis for longitudinal data," *Ann. Stat.* **33**, 2873–2903 (2005).
31. J. Gertheiss et al., "Longitudinal scalar-on-functions regression with application to tractography data," *Biostatistics* **14**(3), 447–461 (2013).
32. J. S. Morris, "Functional regression," *Annu. Rev. Stat. Appl.* **2**(1), 321–359 (2015).
33. M. W. Weiner et al., "The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception," *Alzheimer's Dementia* **9**(5), e111–e194 (2013).
34. C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imaging* **27**(4), 685–691 (2008).
35. S. G. Mueller et al., "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI)," *Alzheimer's Dementia* **1**, 55–66 (2005).
36. S. G. Mueller et al., "The Alzheimer's disease neuroimaging initiative," *Neuroimaging Clin.* **15**(4), 869–877 (2005).
37. N. Chow et al., "Comparing 3T and 1.5T MRI for mapping hippocampal atrophy in the Alzheimer's disease neuroimaging initiative," *Am. J. Neuroradiol.* **36**(4), 653–660 (2015).
38. R. S. Desikan et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage* **31**(3), 968–980 (2006).
39. B. Fischl et al., "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron* **33**, 341–355 (2002).
40. B. Fischl, "Automatically parcellating the human cerebral cortex," *Cereb. Cortex* **14**, 11–22 (2004).
41. D. Ma et al., "Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database," *Hum. Brain Mapp.* **40**(5), 1507–1527 (2018).
42. L. E. Collij et al., "Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease," *Radiology* **281**(3), 865–875 (2016).
43. P. C. O'Brien and P. J. Dyck, "Procedures for setting normal values," *Neurology* **45**(1), 17–23 (1995).
44. R. La Joie et al., "Region-specific hierarchy between atrophy, hypometabolism, and -amyloid (A) load in Alzheimer's disease dementia," *J. Neurosci.* **32**(46), 16265–16273 (2012).
45. J. Young et al., "Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment," *Neuroimage Clin.* **2**, 735–745 (2013).
46. D. Ma et al., "Study the longitudinal *in vivo* and cross-sectional *ex vivo* brain volume difference for disease progression and treatment effect on mouse model of tauopathy using automated MRI structural parcellation," *Front. Neurosci.* **13**, 11 (2019).
47. J. J. Dong et al., "Functional principal component analysis of glomerular filtration rate curves after kidney transplant," *Stat. Methods Med. Res.* **27**(12), 3785–3796 (2018).
48. C. X. Feng, J. Cao, and L. Bendell, "Exploring spatial and temporal variations of cadmium concentrations in pacific oysters from British Columbia," *Biometrics* **67**(3), 1142–1152 (2011).

49. W. Luo et al., "Estimating the intensity of ward admission and its effect on emergency department access block," *Stat. Med.* **32**(15), 2681–2694 (2013).
50. W. Jones and A. J. Klin, "Attention to eyes is present but in decline in 2–6 month-olds later diagnosed with autism," *Nature* **504**(7480), 427–431 (2013).
51. H.-G. Müller, R. Sen, and U. Stadtmüller, "Functional data analysis for volatility," *J. Econometr.* **165**(2), 233–245 (2011).
52. K. Popuri et al., "Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging," *Neuroimage Clin.* **18**, 802–813 (2018).
53. N. Bhagwat et al., "Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data," *PLoS Comput. Biol.* **14**(9), e1006376 (2018).
54. R. Cui and M. Liu, "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease," *Comput. Med. Imaging Graph.* **73**, 1–10 (2019).

Biographies of the authors are not available.