Orthogonal Mixed-Effects Modeling for High-Dimensional Longitudinal Data: An Unsupervised Learning Approach

Ming Chen, Yijun Bian, Nanguang Chen, and Anqi Qiu, *Senior Member, IEEE*

***Abstract*— The linear mixed-effects model is commonly utilized to interpret longitudinal data, characterizing both the global longitudinal trajectory across all observations and longitudinal trajectories within individuals. However, characterizing these trajectories in high-dimensional longitudinal data presents a challenge. To address this, our study proposes a novel approach, Unsupervised Orthogonal Mixed-Effects Trajectory Modeling (UOMETM), that leverages unsupervised learning to generate latent representations of both global and individual trajectories. We design an autoencoder with a latent space where an orthogonal constraint is imposed to separate the space of global trajectories from individual trajectories. We also devise a cross-reconstruction loss to ensure consistency of global trajectories and enhance the orthogonality between representation spaces. To evaluate UOMETM, we conducted simulation experiments on images to verify that every component functions as intended. Furthermore, we evaluated its performance and robustness using longitudinal brain cortical thickness from two Alzheimer's disease (AD) datasets. Comparative analyses with state-of-the-art methods revealed UOMETM's superiority in identifying global and individual longitudinal patterns, achieving a lower reconstruction error, superior orthogonality, and higher accuracy in AD classification and conversion forecasting. Remarkably, we found that the space of global trajectories did not significantly contribute to AD classification compared to the space of individual trajectories, emphasizing their clear separation. Moreover, our model exhibited satisfactory generalization and robustness across different datasets. The study shows the outstanding performance and potential clinical use of UOMETM in the context of longitudinal data analysis.**

***Index Terms*— Longitudinal trajectory, mixed-effects model, orthogonal representation, unsupervised learning.**

I. Introduction

LONGITUDINAL neuroimaging studies have been conducted to understand brain development and aging, the progress of neurodevelopmental disorders and neurodegenerative diseases [1], [2], [3], [4]. Specifically, longitudinal observations of a subject acquired at multiple visits can manifest a longitudinal trajectory that captures temporal changes in structural or functional characteristics of the brain, and provides a quantitative and explicit representation of neurological diseases. Two significant features in modeling longitudinal data are: the average measurement trajectory over time (aka. global trajectory) and the correlation structure among serial measurements (aka. individual heterogeneity).

For low-dimensional longitudinal data, many methods have been proposed and widely applied, including traditional statistical methods (e.g., within-subject ANOVA [5], [6]) and cross-sectional analysis of summary measurements (e.g., assessing annualized differences [7]). However, they fall short in accurately modeling the covariance structure of serial measurements, and struggle with irregular timing or subject dropout, such as unbalanced data. In contrast, linear mixed-effects (LME) models can impose structure on the covariance through the introduction of random effects, and are particularly well suited to handling longitudinal data that are irregularly timed. LME models have widely been employed to analyze longitudinal data for their ability to capture a global progression trajectory expressing group-level characteristics and individual progression trajectories that account for inter-subject variability and diversity [8], both of which are crucial features in longitudinal neuroimaging studies.

Nevertheless, employing LME on high-dimensional data is not straightforward. Researchers have sought to maximize the effectiveness of LME models by integrating additional knowledge to adapt them for handling the complexities of

Manuscript received 18 May 2024; accepted 24 July 2024. Date of publication 30 July 2024; date of current version 2 January 2025. This work was supported in part by the National Science and Technology Innovation 2030 (STI 2030)-Major Project under Grant 2022ZD0209000 and in part by the Hong Kong General Research Fund under Grant 15201124. *(Corresponding author: Anqi Qiu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Central Institutional Review Board (CIRB) at the University of California, San Diego for the ABCD Research and the IRB of Washington University School of Medicine for the OASIS-3 Projects.

Ming Chen, Yijun Bian, and Nanguang Chen are with the Department of Biomedical Engineering, National University of Singapore, Singapore 117583 (e-mail: mchen88@u.nus.edu; yjbian@nus.edu.sg; biecng@nus.edu.sg).

Anqi Qiu is with the Department of Biomedical Engineering, National University of Singapore, Singapore 117583, also with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, and also with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: an-qi.qiu@polyu.edu.hk).

Digital Object Identifier 10.1109/TMI.2024.3435855

processing high-dimensional data. For example, recent studies have enhanced the power of LME models by integrating them with geometric methods, assuming that observations (e.g., 3D images) follow a curve on a Riemannian manifold translated from a common geodesic, allowing effective parametrization of space shift and time variation [9], [10], [11], [12], [13]. However, it is not straightforward to design a Riemannian space with a metric that is suitable for the distribution of high-dimensional longitudinal data.

Recently, several approaches combining deep learning techniques—especially in an unsupervised manner—have been proposed to extract informative low-dimensional representations [11], [14], [15], [16], [17]. Ouyang et al. [17] employed an autoencoder to project high-dimensional image data into a latent space associated with brain age. However, without purposeful constraints on the latent space, unsupervised learning is likely to generate representations with weak relations to interpretable real-world factors like age, gender, and disease diagnosis, thus leading to ambiguous results. To enrich interpretability, learning disentangled representations of generative factors within data emerges [18], [19], [20], [21], [22], [23]. Through disentanglement, it is separatable between within-subject and group-level variability when fitting longitudinal trajectories, thereby enhancing the interpretability of generated low-dimensional representations [24]. For this, Higgins et al. [25] achieved disentanglement by adding a hyperparameter on a variational autoencoder (VAE) to emphasize learning statistically independent latent factors. Multi-level VAE (ML-VAE) proposed by Bouchacourt et al. [26] learned disentangled representations of a set of group observations and separated latent representations into content and style. Couronne et al. [15] disentangled longitudinal observations into two representation spaces, each associated with inter-patient variability and time progression variability respectively, offering explicit practical interpretations. Nevertheless, these existing approaches do not guarantee that the disentangled representations correspond to the desired global or individual trajectories, because they may lack the ability to completely eliminate the potential overlap or intersection between disentangled representations. Hence, there is a necessity for improvements to strictly enforce independence or perpendicularity among disentangled representation spaces, ensuring more precise and reliable interpretations of the underlying data characteristics.

To address these issues, we propose *Unsupervised Orthogonal Mixed-Effects Trajectory Modeling (UOMETM)*, leveraging the unsupervised nature of an autoencoder framework. In the latent space, we generate orthogonal representations and construct a parametric mixed-effects progression model on these representations to flexibly and integratively model global and individual longitudinal trajectories. Our novel orthogonal constraint significantly establishes the perpendicularity between representation spaces for global and individual longitudinal trajectories, thus enabling unbiased and reliable interpretations. Our contributions in this work are four-fold:

- We propose a representation learning method named *UOMETM* for longitudinal trajectory modeling, which is able to extract low-dimensional representations from high-dimensional data in an unsupervised manner.
- We perform an orthogonal constraint on representations to ensure strict perpendicularity between the spaces of global and individual trajectories for interpretability.
- We formulate a parametric progression model based on a linear mixed-effects model to flexibly capture global and individual trajectories, while constructing its nested models on the orthogonal representations.
- We validate the mechanisms of each component of *UOMETM* on a simulation dataset and then evaluate the whole model performance on two Alzheimer's disease (AD) datasets, using several fundamental metrics, comprehensive assessments, and clinical downstream tasks. We further verify the generalization and robustness of *UOMETM* across different datasets.

II. Related Works

A. High-Dimensional Longitudinal Studies With LME Models

Since LME models are flexible and parsimonious for modeling covariance among observations, especially adept at accommodating longitudinal data with irregular time intervals [8], they are widely used for longitudinal trajectory modeling [10], [12], [27], [28], [29]. Yet for high-dimensional data, some challenges remain even though LME models have demonstrated remarkable performance in low-dimensional longitudinal studies. This is primarily due to the complex covariance structures inherent in high-dimensional data, which complicate longitudinal analysis.

To overcome this problem, several methods have incorporated geometric approaches to enhance the applicability of LME models [9], [10], [11], [12], [13]. For example, Sauty and Durrleman [29] proposed Riemannian VAE (Riem-VAE) which utilized a temporal linear mixed-effect model to learn latent representations which aligned with expected geodesics on a Riemannian manifold. Similarly, Schiratti et al. [10] proposed a Bayesian mixed-effects model to learn spatio-temporal trajectories from manifold-valued longitudinal data. Among these geometric strategies, establishing the metric for the Riemannian manifold poses significant challenges. Recently, unsupervised learning, particularly via autoencoders, has proven effective in managing unlabeled high-dimensional data while offering flexibility in imposing constraints within the latent space. For instance, Ouyang et al. [17] introduced Longitudinal Neighbourhood Embedding (LNE), employing an autoencoder to produce low-dimensional latent representations while employing a contrastive loss on the latent space. Consequently, our study opts for an autoencoder framework, to generate informative low-dimensional representations from high-dimensional image or feature vector data. These representations can effectively facilitate the utility of LME models.

B. Representation Disentanglement in Longitudinal Studies

Recently, representation disentanglement has been widely applied in various applications [18], [19], [20], [21], [22],

[23]. A disentangled representation is defined as a single latent unit that is only related to variations in one single generative factor and invariant to changes in other factors [18], [25]. This makes disentangled representations highly interpretable as the underlying generative factors have strong associations with real-world factors, such as age and disease diagnosis [30].

In longitudinal studies, disentanglement becomes particularly advantageous as it allows for the untangling of underlying factors of variation within the data, enabling the investigation of global trajectory and individual trajectory separately. Consequently, representation disentanglement finds widespread application in longitudinal neuroimaging studies. For instance, Higgins et al. [25] proposed β-VAE which leveraged a hyperparameter β to discover interpretable factorized latent representations from high-dimensional data. ML-VAE introduced by Bouchacourt et al. [26] separated representations into content and style spaces. To enhance the interpretability of disentangled representation spaces, Couronne et al. [15] proposed Rank-VAE to encode input images into inter-patient variability and time progression variability, and further add a ranking loss which forced the time progression variability to present a preferential progression order. However, a notable challenge arises in ensuring clear separation between representation spaces, as potential information intersection or leakage may occur. Such an issue could lead to inaccurate interpretations of disentangled representation spaces. To address this challenge, we propose imposing a constraint on the representation spaces to ensure orthogonality between them. This constraint aims to guarantee thorough separation between the space of global trajectory and the space of individual trajectory, thereby facilitating accurate investigations of global patterns and individual heterogeneity within longitudinal observations.

III. Methods

In this section, we elaborate on the proposed *Unsupervised Orthogonal Mixed-Effects Trajectory Modeling* (*UOMETM*). Our *UOMETM* framework comprises of three components: an encoder $\Phi(\cdot)$, orthogonal mixed effects trajectory modeling (*OMETM*) in a latent space, and a decoder $\Psi(\cdot)$, as illustrated in Fig. 1(a). In the following, we describe in detail the methodology of *OMETM* and regularization that are necessary to produce consistent and interpretable global and individual longitudinal trajectories. Note that the encoder and decoder in this study are implemented as cascaded convolutional layers when processing image data or as a series of fully connected layers when inputs are feature vectors.

A. Notations

For readability, we begin by introducing the essential notations used in this paper. We use italic lowercase letters (e.g., x), bold lowercase letters (e.g., $\mathbf{x}$), and bold uppercase letters (e.g., $\mathbf{X}$) to represent scalars, vectors/tensors, and matrices, respectively. Note that $\mathbb{R}$ is used to denote a real space, and $i \in [n]$ represents $i \in \{1, 2, \ldots, n\}$ for brevity. Notations $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{MN}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ stand for Gaussian distribution and matrix Gaussian distribution, respectively. Operator symbol $\|\cdot\|$ denotes the Euclidean norm.

In this paper, we have a set of high-dimensional observations from m subjects in total, that is, $\{\mathbf{x}_{ij} \mid j \in [n_i],\ i \in [m]\}$, of which each high-dimensional observation/sample $\mathbf{x}_{ij}$ is observed as the j-th visit at time t_{ij}, where n_i indicates the number of visits of the i-th subject, and $\sum_{i=1}^{m} n_i = N$ represents the number of observations/samples in total. Note that symbols $\mathbf{0}$ and $\mathbf{I}$ are used to denote a zero matrix and an identity matrix, respectively.

B. Orthogonal Mixed-Effects Trajectory Modeling

We extract low-dimensional representations by implementing an encoder network $\Phi(\cdot)$ from an autoencoder framework to map input data $\mathbf{x}_{ij}$ to a latent space, that is,

$$\mathbf{z}_{ij} = \Phi(\mathbf{x}_{ij}) \in \mathbb{R}^L\,,\ \forall\, j \in [n_i],\ i \in [m]\,, \tag{1}$$

where L is the dimension of the latent space.

1) A Mixed-effects Model: Analogous to a traditional linear mixed-effects model, we formulate the latent space to be a mixed-effects model, which is flexible to express global characteristics and individual heterogeneity of longitudinal progression. We define $\mathbf{t}_{ij}^{\text{fe}} = [1, t_{ij} - t_{i1}, t_{i1}]^{\mathsf{T}} \in \mathbb{R}^3$ and $\mathbf{t}_{ij}^{\text{re}} = [1, t_{ij} - t_{i1}]^{\mathsf{T}} \in \mathbb{R}^2$ that will constitute design matrices for fixed effects and random effects, respectively. We let $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3] \in \mathbb{R}^{L\times 3}$ and $\mathbf{B}_i = [\mathbf{b}_{i1}, \mathbf{b}_{i2}] \in \mathbb{R}^{L\times 2}$ denote fixed effect and random effect coefficients, respectively. Then, $\boldsymbol{\beta}\mathbf{t}_{ij}^{\text{fe}}$ would be the fixed effects capturing a global trajectory that is supposed to summarize group-level characteristics of the longitudinal progression. The latent representations, observations $\{\mathbf{z}_{ij} \mid j \in [n_i], i \in [m]\}$ of the i-th subject at the j-th visit, can be modeled as

$$\begin{aligned}\mathbf{z}_{ij} &= \boldsymbol{\beta}\mathbf{t}_{ij}^{\text{fe}} + \mathbf{B}_i\mathbf{t}_{ij}^{\text{re}} + \boldsymbol{\epsilon}_{ij}^{0}\\ &= \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \cdot (t_{ij} - t_{i1}) + \boldsymbol{\beta}_3 \cdot t_{i1}\\ &\quad + \mathbf{b}_{i1} + \mathbf{b}_{i2} \cdot (t_{ij} - t_{i1}) + \boldsymbol{\epsilon}_{ij}^{0},\end{aligned} \tag{2}$$

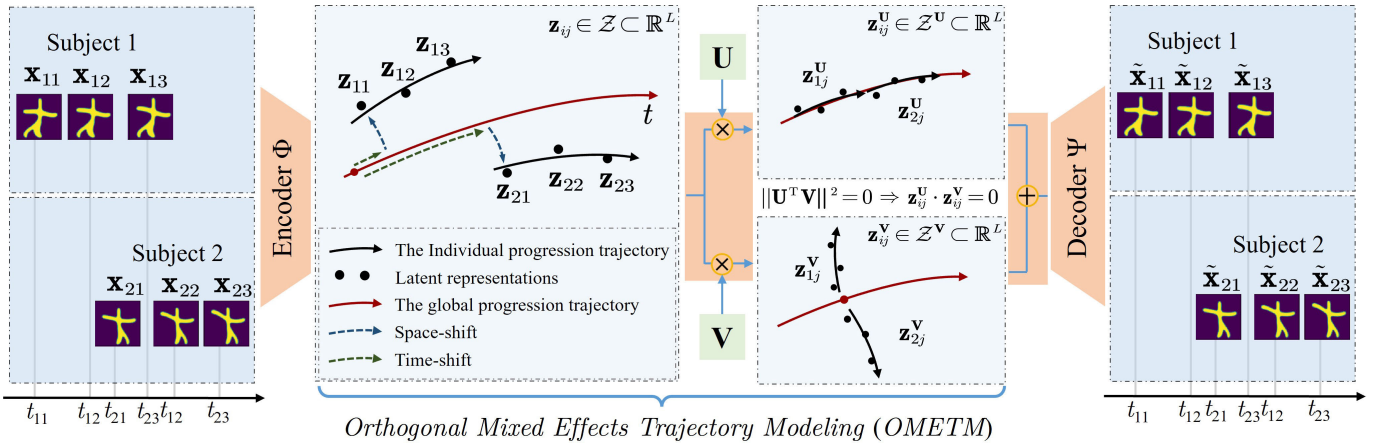
where $\boldsymbol{\epsilon}_{ij}^{0} \sim \mathcal{N}(0, \sigma_0^2\mathbf{I})$ is noise, independent and identically distributed (i.i.d) from each other. All vectors (i.e., $\mathbf{z}_{ij}$, $\boldsymbol{\beta}_k$ for $k \in [3]$, $\mathbf{b}_{ik}$ for $k \in [2]$, and $\boldsymbol{\epsilon}_{ij}^{0}$) belong to the space $\mathbb{R}^L$. Note that $\mathbf{b}_{i1}$ encodes the spatial shift of individual progression trajectories; and $\mathbf{b}_{i2} \cdot (t_{ij} - t_{i1})$ is the formulation of the time-parametrizing strategy to characterize temporal heterogeneity, where the time-shift t_{i1} is the onset and the acceleration factor $\mathbf{b}_{i2}$ accounts for the speed of longitudinal progression.

2) Orthognoality: In this study, we aim to best separate individual progression trajectories from the global trajectory. For this, we define two linear transformations (that is, $\mathbf{U}$ and $\mathbf{V} \in \mathbb{R}^{L\times L}$) such that

$$\mathbf{z}_{ij}^{\mathbf{U}} \triangleq \mathbf{U}\mathbf{z}_{ij} = \boldsymbol{\beta}\mathbf{t}_{ij}^{\text{fe}} + \boldsymbol{\epsilon}_{ij}^{1}, \tag{3a}$$

$$\mathbf{z}_{ij}^{\mathbf{V}} \triangleq \mathbf{V}\mathbf{z}_{ij} = \mathbf{B}_i\mathbf{t}_{ij}^{\text{re}} + \boldsymbol{\epsilon}_{ij}^{2}, \tag{3b}$$

where $\boldsymbol{\epsilon}_{ij}^{k} \sim \mathcal{N}(0, \sigma_k^2\mathbf{I})$ for $k = 1, 2$ are noise, and $\boldsymbol{\epsilon}_{ij}^{k} \in \mathbb{R}^L$. Here, we impose an orthogonality property between $\mathbf{z}_{ij}^{\mathbf{U}}$ and $\mathbf{z}_{ij}^{\mathbf{V}}$ for their best separation, which is achieved by introducing



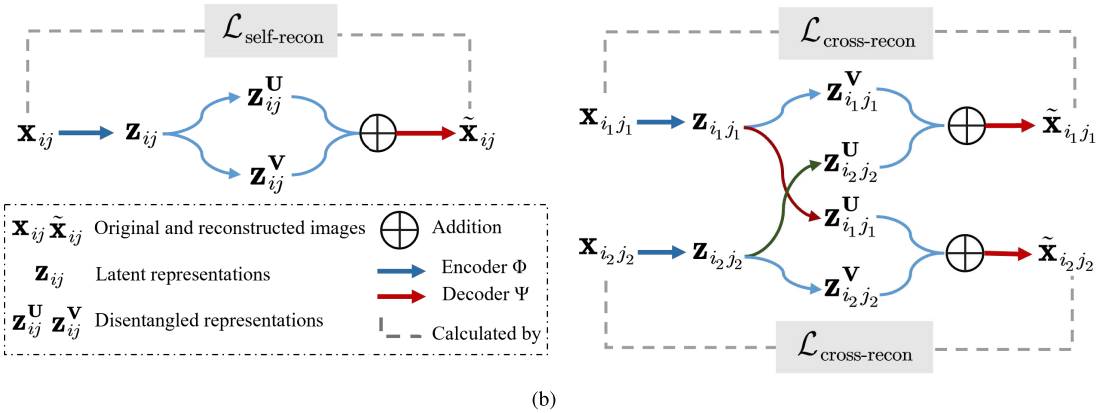


Fig. 1. Schematic of the proposed *UOMETM* framework. (a) The input data $\mathbf{x}_{ij}$ undergoes encoding into latent representations $\mathbf{z}_{ij}$. Subsequently, orthogonal mixed effects trajectory modeling (*OMETM*) is constructed in the latent space to capture accurate global ($\mathbf{z}_{ij}^{\mathbf{U}}$) and individual trajectories ($\mathbf{z}_{ij}^{\mathbf{V}}$), leveraging the orthogonality through linear transformations $\mathbf{U}$ and $\mathbf{V}$. Finally, the orthogonal representations are incorporated into a decoder to yield the reconstructed outcomes $\tilde{\mathbf{x}}_{ij}$. (b) The left panel shows the self-reconstruction loss that is quantified as the mean squared error (MSE) between the input and the reconstructed output when decoding the orthogonal representations from the same sample. The right panel illustrates the cross-reconstruction loss that ensures the consistency of global trajectories among all sample pairs from $\mathcal{S}$, each pair containing two samples with the same age.

a constraint on $\mathbf{U}$ and $\mathbf{V}$, that is, $\mathbf{U}^{\mathsf{T}}\mathbf{V} = \mathbf{0}$. As a result, the inner product of $\mathbf{z}_{ij}^{\mathbf{U}}$ and $\mathbf{z}_{ij}^{\mathbf{V}}$ can be written as

$$(\mathbf{z}_{ij}^{\mathbf{U}})^{\mathsf{T}}\mathbf{z}_{ij}^{\mathbf{V}} = (\mathbf{U}\mathbf{z}_{ij})^{\mathsf{T}}(\mathbf{V}\mathbf{z}_{ij}) = \mathbf{z}_{ij}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{V}\mathbf{z}_{ij} = \mathbf{z}_{ij}^{\mathsf{T}}\mathbf{0}\mathbf{z}_{ij} = 0.$$

Hence, we introduce an orthogonal loss, that is,

$$\mathcal{L}_{\text{ortho}} = \|\mathbf{U}^{\mathsf{T}}\mathbf{V}\|^2, \tag{4}$$

to ensure $\mathbf{U}^{\mathsf{T}}\mathbf{V} = 0$.

3) OMETM: To combine Equations (2), (3a), and (3b), we define a modeling loss to enforce the progression model and its nested models to comply with the corresponding representations. For brevity, we eliminate subscripts i and j through vector concatenation, and obtain the modeling loss in a matrix form, that is,

$$\begin{aligned}\mathcal{L}_{\text{model}} = \; & \frac{1}{\sigma_0^2}\|\mathbf{Z} - \boldsymbol{\beta}\mathbf{T}_{\text{fe}} - \mathbf{B}\mathbf{T}_{\text{re}}\|^2 + \text{tr}(\mathbf{B}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{B}) \\ & + \frac{1}{\sigma_1^2}\|\mathbf{Z}^{\mathbf{U}} - \boldsymbol{\beta}\mathbf{T}_{\text{fe}}\|^2 + \frac{1}{\sigma_2^2}\|\mathbf{Z}^{\mathbf{V}} - \mathbf{B}\mathbf{T}_{\text{re}}\|^2, \end{aligned} \tag{5}$$

such that

$$\begin{aligned}\mathbf{Z} \; &= \boldsymbol{\beta}\mathbf{T}_{\text{fe}} + \mathbf{B}\mathbf{T}_{\text{re}} + \boldsymbol{\epsilon}_0, \\ \mathbf{Z}^{\mathbf{U}} &= \mathbf{U}\mathbf{Z} = \boldsymbol{\beta}\mathbf{T}_{\text{fe}} + \boldsymbol{\epsilon}_1, \\ \mathbf{Z}^{\mathbf{V}} &= \mathbf{V}\mathbf{Z} = \mathbf{B}\mathbf{T}_{\text{re}} + \boldsymbol{\epsilon}_2,\end{aligned}$$

where $\mathbf{B} \sim \mathcal{MN}(\mathbf{0}, \mathbf{D}, \mathbf{I})$ and $\mathbf{D}$ is an $L \times L$ covariance matrix. Here, coefficients $\boldsymbol{\beta}$ and $\mathbf{B}$, and the variance parameters denoted as $\theta = (\sigma_0^2, \sigma_1^2, \sigma_2^2, \mathbf{D})$ need to be estimated. Through minimizing Eq. (5), our model can establish a global progression trajectory for the entire population and individual progression trajectories for each subject.

C. Self- and Cross-Reconstruction

We design the *UOMETM* framework as an unsupervised method by introducing a decoder $\Psi(\cdot)$ that can recover observations from the above latent space. The customized reconstruction losses are proposed here for two purposes: one is to optimize the autoencoder network parameters, and the other is to ensure the consistency of the global trajectory across samples.

First, we define the mean squared error (MSE) between input $\mathbf{x}_{ij}$ and its reconstruction, $\tilde{\mathbf{x}}_{ij} = \Psi(\mathbf{z}_{ij}^{\mathbf{U}} + \mathbf{z}_{ij}^{\mathbf{V}})$, as a self-reconstruction loss. This loss serves as a conventional

regularization for such an autoencoder framework, that is,

$$\mathcal{L}_{\text{self-recon}} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\|\mathbf{x}_{ij} - \Psi(\mathbf{z}_{ij}^{\mathbf{U}} + \mathbf{z}_{ij}^{\mathbf{V}})\right\|^2. \tag{6}$$

Second, in the *UOMETM* framework, $\mathbf{z}_{ij}^{\mathbf{U}}$ is designed to be common among samples with the same age and $\mathbf{z}_{ij}^{\mathbf{V}}$ encodes individual heterogeneity. For a pair of two observations $(i_1 j_1, i_2 j_2) \in \mathcal{S}$, the reconstruction $\Psi(\mathbf{z}_{i_2 j_2}^{\mathbf{U}} + \mathbf{z}_{i_1 j_1}^{\mathbf{V}})$ is supposed to be similar to $\mathbf{x}_{i_1 j_1}$. Symmetrically, $\Psi(\mathbf{z}_{i_1 j_1}^{\mathbf{U}} + \mathbf{z}_{i_2 j_2}^{\mathbf{V}})$ can approximate $\mathbf{x}_{i_2 j_2}$. In other words,

$$\mathbf{x}_{i_1 j_1} \approx \Psi\left(\mathbf{z}_{i_2 j_2}^{\mathbf{U}} + \mathbf{z}_{i_1 j_1}^{\mathbf{V}}\right), \tag{7a}$$

$$\mathbf{x}_{i_2 j_2} \approx \Psi\left(\mathbf{z}_{i_1 j_1}^{\mathbf{U}} + \mathbf{z}_{i_2 j_2}^{\mathbf{V}}\right). \tag{7b}$$

To combine Equations (7a) and (7b), we define a cross-reconstruction loss. The so-called "cross" indicates that the orthogonal representations to be decoded come from different samples, that is,

$$\mathcal{L}_{\text{cross-recon}} = \frac{1}{2|\mathcal{S}|}\sum_{(i_1 j_1, i_2 j_2)\in\mathcal{S}}\left(\left\|\mathbf{x}_{i_1 j_1} - \Psi\left(\mathbf{z}_{i_2 j_2}^{\mathbf{U}} + \mathbf{z}_{i_2 j_2}^{\mathbf{V}}\right)\right\|^2 + \left\|\mathbf{x}_{i_2 j_2} - \Psi\left(\mathbf{z}_{i_1 j_1}^{\mathbf{U}} + \mathbf{z}_{i_2 j_2}^{\mathbf{V}}\right)\right\|^2\right), \tag{8}$$

where $\mathcal{S}$ is a set containing pairs of samples with the same age, each of which expresses a consistent global progression. Note that $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. By incorporating the cross-reconstruction loss into the self-reconstruction loss during the training, we are able to enhance the robustness of the orthogonal representations and eliminate the potential overlap that might occur among them.

The performance of $\mathcal{L}_{\text{cross-recon}}$ can be evaluated through the consistency of the global trajectory, which is calculated by

$$\frac{1}{|\mathcal{S}|\cdot \text{nop}}\sum_{(i_1 j_1, i_2 j_2)\in\mathcal{S}}\left\|\Psi(\mathbf{z}_{i_1 j_1}^{\mathbf{U}}) - \Psi(\mathbf{z}_{i_2 j_2}^{\mathbf{U}})\right\|^2, \tag{9}$$

where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$, and nop represents the number of pixels of the reconstructed outcome of $\Psi(\cdot)$.

D. Loss Functions and Optimization

Our study involves estimating parameters in three main parts: i) the network parameters for encoder Φ and decoder Ψ; ii) linear transformations of $\mathbf{U}$ and $\mathbf{V}$; and iii) coefficients $\boldsymbol{\beta}$ and $\mathbf{B}$, and variance parameters θ for the global and individual trajectories. We partition these parameters into two blocks and formulate two overall loss functions, each corresponding to a block: $\text{blk}_1 = \{\Phi, \Psi\}$ and $\text{blk}_2 = \{\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta\}$. An iterative optimization algorithm, incorporated with the block coordinate descent (BCD) method, is proposed and described in Algorithm 1 for optimizing these loss functions and addressing complex interdependence among parameters. Through our BCD-based optimization, we iteratively optimize each block while freezing the other to find the optimal solution.

To optimize the parameters in blk_1, we define a loss that consists of the self-recon loss in Eq. (6) and the cross-recon loss in Eq. (8), that is,

$$\mathcal{L}(\Phi, \Psi \mid \mathbf{U}, \mathbf{V}) = \mathcal{L}_{\text{self-recon}} + \mathcal{L}_{\text{cross-recon}}. \tag{10}$$

The optimization of Φ and Ψ can be achieved via a backpropagation algorithm. When calculating $\mathcal{L}_{\text{cross-recon}}$, since we establish $\mathcal{S}$ by conducting pairwise comparisons among all pairs of observations, the computational complexity is $\mathcal{O}(N^2)$, which is computationally intensive. Therefore, we determine $\mathcal{S}$ for each batch instead of the entire dataset in every iteration, resulting in $\mathcal{O}(\text{bs}^2)$ complexity, where bs is the batch size.

About the parameters in blk_2, their optimizations are incorporated into a loss function involving the modeling loss in Eq. (5) and the orthogonal loss in Eq. (4) given Φ to compute the orthogonal representations, that is,

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi) = \mathcal{L}_{\text{model}} + \tfrac{\lambda}{2}\mathcal{L}_{\text{ortho}}, \tag{11}$$

where a hyperparameter λ is used to control relative contributions between these two terms. We can derive the analytic solutions for these parameters by letting the partial derivative with respect to (w.r.t.) each of them equal to zero, that is,

$$\frac{\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)}{\partial\boldsymbol{\beta}} = 0, \tag{12a}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)}{\partial\mathbf{B}} = 0, \tag{12b}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)}{\partial\mathbf{U}} = 0, \tag{12c}$$

$$\frac{\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)}{\partial\mathbf{V}} = 0. \tag{12d}$$

Algorithm 1 BCD-Based Optimization

Input: Observations $\{\mathbf{x}_{ij} \mid j \in [n_i], i \in [m]\}$, hyperparameters λ, and L

Output: Φ, Ψ, $\mathbf{U}$, $\mathbf{V}$, $\boldsymbol{\beta}$, $\mathbf{B}$, and θ

Initialize $\mathbf{U} := \text{diag}(1 \text{ or } 0)$ with $\lfloor \frac{L}{2} \rfloor$ zero diagonal elements
Initialize $\mathbf{V} := \mathbf{I} - \mathbf{U}$
for each epoch **do**
  % Update $\text{blk}_1 = \{\Phi, \Psi\}$ *with fixed* blk_2
  Compute $\mathcal{L}_{\text{self-recon}}$ and $\mathcal{L}_{\text{cross-recon}}$
  $\mathcal{L}(\Phi, \Psi \mid \mathbf{U}, \mathbf{V}) = \mathcal{L}_{\text{self-recon}} + \mathcal{L}_{\text{cross-recon}}$
  Update Φ and Ψ by backpropagating $\mathcal{L}(\Phi, \Psi \mid \mathbf{U}, \mathbf{V})$
  % Update $\text{blk}_2 = \{\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta\}$ *with fixed* blk_1
  repeat
    Compute $\mathcal{L}_{\text{model}}$ and $\mathcal{L}_{\text{ortho}}$
    $\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi) = \mathcal{L}_{\text{model}} + \frac{\lambda}{2}\mathcal{L}_{\text{ortho}}$
    Compute $\boldsymbol{\beta}$ by letting $\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)/\partial\boldsymbol{\beta} = 0$
    Compute $\mathbf{B}$ by letting $\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)/\partial\mathbf{B} = 0$
    Compute $\mathbf{U}$ by letting $\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)/\partial\mathbf{U} = 0$
    Compute $\mathbf{V}$ by letting $\partial\mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)/\partial\mathbf{V} = 0$
  until convergence
  $\theta := \text{argmin}_\theta\, \mathcal{L}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{U}, \mathbf{V}, \theta \mid \Phi)$
end for

Hence, we obtain that

$$\boldsymbol{\beta} = \frac{1}{\sigma_0^2 + \sigma_1^2}(\sigma_1^2(\mathbf{Z} - \mathbf{B}\mathbf{T}_{\text{re}})\mathbf{T}_{\text{fe}}^{\mathsf{T}} + \sigma_0^2\mathbf{Z}^{\mathbf{U}}\mathbf{T}_{\text{fe}}^{\mathsf{T}})(\mathbf{T}_{\text{fe}}\mathbf{T}_{\text{fe}}^{\mathsf{T}})^{-1}, \tag{13a}$$

$$\mathbf{B} = \left(\sigma_2^2(\mathbf{Z} - \boldsymbol{\beta}\mathbf{T}_{\text{fe}})\mathbf{T}_{\text{re}}^{\mathsf{T}} + \sigma_0^2\mathbf{Z}^{\mathbf{V}}\mathbf{T}_{\text{re}}^{\mathsf{T}}\right) \cdot \left((\sigma_0^2 + \sigma_2^2)\mathbf{T}_{\text{re}}\mathbf{T}_{\text{re}}^{\mathsf{T}} - 2\sigma_0^2\sigma_2^2\mathbf{D}^{-1}\right)^{-1}, \tag{13b}$$

$$\mathbf{U} = (\mathbf{Z}\mathbf{Z}^{\mathsf{T}} + \sigma_1^2\lambda\mathbf{V}\mathbf{V}^{\mathsf{T}})^{-1}\mathbf{Z}(\boldsymbol{\beta}\mathbf{T}_{\text{fe}})^{\mathsf{T}}, \tag{13c}$$

$$\mathbf{V} = (\mathbf{Z}\mathbf{Z}^{\mathsf{T}} + \sigma_2^2\lambda\mathbf{U}\mathbf{U}^{\mathsf{T}})^{-1}\mathbf{Z}(\mathbf{B}\mathbf{T}_{\text{re}})^{\mathsf{T}}. \tag{13d}$$

As for variance parameters $\theta = (\sigma_0^2, \sigma_1^2, \sigma_2^2, \mathbf{D})$, they can be estimated via maximum likelihood with detailed equations explained in Gumedze and Dunne's work [31].

IV. Experimental Setup

In this section, we introduce state-of-the-art (SOTA) baseline methods for model comparison and datasets employed in this study along with their implementation details. The source code is available at https://github.com/MChen808/UOMETM.

A. Baseline Methods

We chose several SOTA models as baseline methods, including β-VAE [25], ML-VAE [26], Rank-VAE [15], Riem-VAE [29], and LNE [17]. They are chosen because they are unsupervised learning approaches for longitudinal trajectory modeling. In particular, ML-VAE, Rank-VAE, and Riem-VAE also model the global and individual trajectories as disentangled components. When conducting extrapolation tasks for unseen observations, we only compared our method with Riem-VAE, the only model capable of predicting future timepoints by leveraging time information among the aforementioned baseline models.

When implementing β-VAE, Riem-VAE, and LNE, each featuring a single encoder and decoder, the network architecture—including the dimensionality of the latent space L, the number of layers, filter sizes, pooling kernel sizes, and activation functions—were set to be identical to our model's configuration. In the case of ML-VAE and Rank-VAE, featuring two encoders for disentangled components and a single decoder for reconstructing concatenated representations, the encoders of ML-VAE were aligned with ours, each projecting to an L-dim latent space. Similarly, the encoders of Rank-VAE mirrored ours, differing only in the dimensions of their respective latent spaces, which were 1-dim and $(L-1)$-dim. The decoder of Rank-VAE conformed to our model's configuration, while that of ML-VAE exhibited a similar structure, distinct solely in the input dimension of its first layer, which was adjusted to $2L$. All VAE-based models utilized an identical hyperparameter to balance the reconstruction loss and the Kullback-Leibler divergence. The hyperparameters governing the influence of customized losses for each model were carefully chosen to optimize their overall performance. These settings were adopted to maintain consistency and comparability among different models.

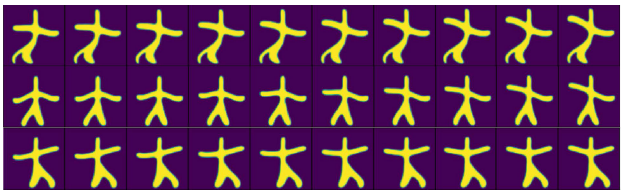


Fig. 2. Examples of synthetic images. Each row shows ten sequential images of one subject.

TABLE I
The Number of Subjects Based on the Number of Visits in Each Diagnostic Group Fromtwo Clinical Datasets. (a) From the ADNI Dataset; (b) From the OASIS-3 Dataset

(a)

Number of visits	CN	s-MCI	AD	Conversion
1	54	45	49	0
2	62	57	47	7
3	52	87	75	13
4	79	127	146	37
5	94	114	7	68
6	31	64	2	59
7	21	28	0	30
8	17	8	0	18
9	18	12	0	12
10	6	6	0	3
11	0	0	0	4
Total	434	548	326	251

(b)

Number of visits	CN	s-MCI	AD	Conversion
1	242	42	187	0
2	191	3	44	8
3	114	2	1	8
4	64	0	1	1
5	26	0	0	0
6	13	0	0	0
Total	650	47	233	17

B. Datasets

1) Simulation Dataset: We generated a synthetic longitudinal dataset of 64×64 Starmen images with pixel values normalized in the range of [0, 1], based on a longitudinal diffeomorphic model [9]. Some examples of the dataset are displayed in Fig. 2. The *global progression trajectory* was the rising stage of the left arm while the *individual heterogeneity* within this dataset was characterized by the location of other limbs.

In total, we generated $m = 1000$ subjects, each with $n_i = 10$ visits. Hence, this dataset contains a total $N = 10000$ observations.

2) Clinical Datasets: Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1] and the Open Access Series of Imaging Studies-3 (OASIS-3).[2] Institutional review boards approved study procedures across participating institutions.

The ADNI dataset included ADNI-1, ADNI-GO and ADNI-2. At each visit, subjects were diagnosed as one of three clinical statuses: cognitive normal (CN), mild cognitive

[1]http://adni.loni.usc.edu
[2]http://oasis-brains.org

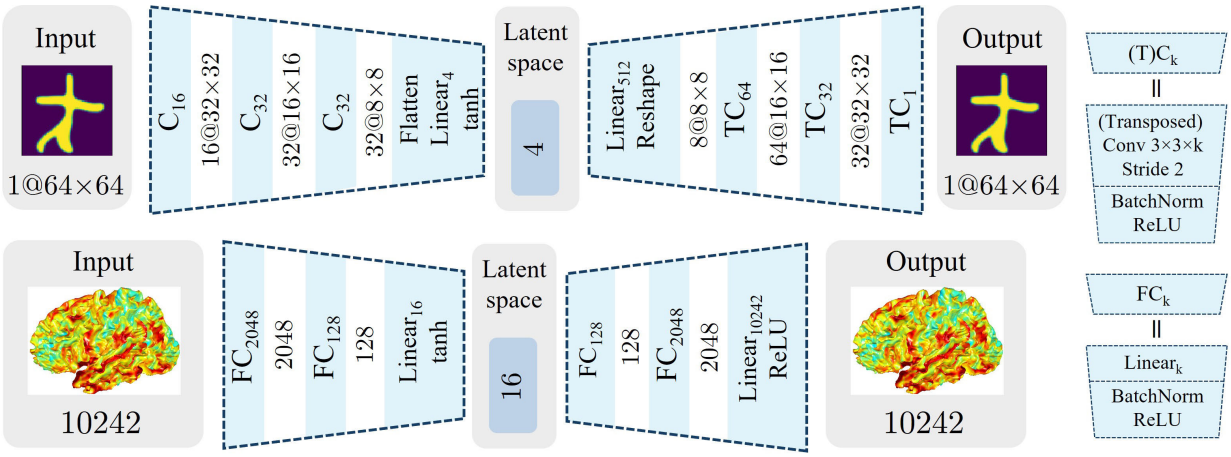


Fig. 3. The autoencoder network architectures for the simulation dataset (top row) and the clinical datasets (bottom row). The right panels respectively show the convolutional block and fully connected block from top to bottom.

impairment (MCI), and Alzheimer's disease (AD) subjects. The number of visits per subject varied from 1 to 11. There were 434 CN subjects, 548 stable MCI (s-MCI) subjects who did not convert to AD across all existing visits, 326 AD subjects, and 251 subjects who converted from CN or MCI to AD. Table I(a) lists the detailed information about the number of subjects based on the number of visits in each diagnostic group.

We also employed the OASIS-3 dataset to validate the generalization and robustness of our model. The number of visits per subject varied from 1 to 6. This study included 650 CN subjects, 47 s-MCI subjects, 233 AD subjects, and 17 subjects converted from CN or MCI to AD. Table I(b) enumerates details regarding the number of subjects in each diagnostic group, sorted by the number of their visits.

The MRI preprocessing for both the ADNI and OASIS-3 datasets is conducted identically. Structural T_1-weighted MRI images were processed using FreeSurfer (version 5.3.0) to extract cortical thickness as brain morphological features. We employed large deformation diffeomorphic metric mapping (LDDMM) [32] to align individual cortical surfaces to the atlas and transfer cortical thickness of each subject to a common space, which was used in the following experiments. For each hemisphere of a subject, we had the cortical thickness data as vectors with 163,842 dimensions, which we firstly downsampled into vectors with only 10,242 dimensions using the fast vertex-based graph convolutional neural network [33]. If any visualization on the cortical surface is required, we will upsample the vectors with 10,242 dimensions to the original 163,842 dimensions through the neighborhood matrices recorded during the downsampling. Hence, the 10,242-dim feature vectors serve as the input data during the clinical experiments, as presented in Fig. 3.

3) Experimental Details: All experiments conducted on the simulation dataset were evaluated via 5-fold cross-validation. The experiments of the ADNI data were conducted for ten times with half of the subjects randomly selected as a training set and the remaining half as a test set. As for the OASIS-3 dataset, we confined our usage to CN and AD subjects. These subjects served as an additional test set for the ADNI dataset, helping verify the robustness of the model we proposed during performing the AD vs. CN classification task.

C. Network Structure

The network structure for two datasets is well designed with details presented in Fig. 3. For the simulation dataset, the autoencoder architecture was constructed with cascaded convolutional layers to process image data. Let C_k denote a Convolution–BatchNorm–ReLU block including k convolutional filters with stride 2 and filter size 3×3, batch normalization, and rectified linear unit (ReLU). TC_k represents a Transposed Conv–BatchNorm–ReLU block. For this dataset, the encoder Φ was designed as C_{16}–C_{32}–C_{32} while the decoder Ψ was TC_{64}–TC_{32}–TC_1. The output of encoding was flattened and projected as 4-dim latent representations through a linear layer and an activation function tanh. Before decoding, the resultant representations were transformed into 512-dim representations via a linear layer and were reshaped as $8@8 \times 8$ tensors.

For the ANDI experiments, the autoencoder architecture was built based on a series of fully-connected layers without convolutional layers, because cortical thickness is not defined in a regular grid (i.e., an image grid). Let FC_k denote a Linear–BatchNorm–ReLU block with k nodes. The encoder Φ was designed as FC_{2048}–FC_{128}–Linear_{16}. The decoder Ψ was designed as FC_{128}–FC_{2048}–Linear_{10242}–ReLU. After encoding, feature vectors were projected into the latent space through an activation function tanh.

D. Hyperparameters

In our proposed model, we introduce two key hyperparameters: L and λ, each crucial for model performance.

The hyperparameter L stands out as particularly critical as it defines the dimensionality of the latent space. In our approach, we conduct mixed-effects modeling on L-dim latent representations while imposing an orthogonal constraint on $L \times L$ matrices $\mathbf{U}$ and $\mathbf{V}$. A smaller value of L significantly reduces computational burden and enhances performance of mixed-effects modeling and orthogonality.

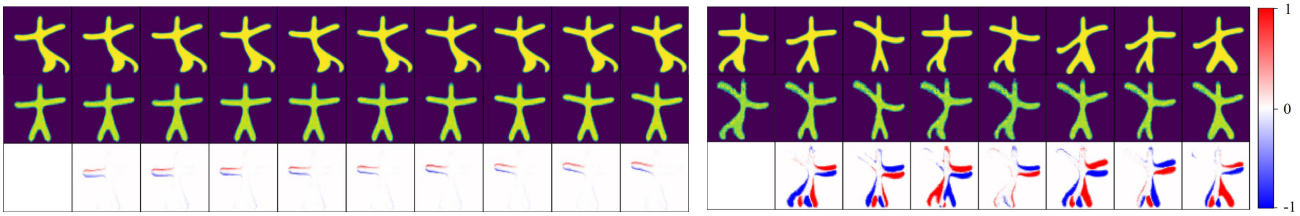


Fig. 4. Visualization of the reconstruction from $\mathcal{Z}^{\mathbf{U}}$ (a) and $\mathcal{Z}^{\mathbf{V}}$ (b). Panel (a) shows an example of the global trajectory. Top to bottom rows respectively show the original images $\mathbf{x}_{ij}$, reconstructed representations from $g(\mathbf{z}_{ij}^{\mathbf{U}})$, and the subtraction between adjacent reconstructions. The subtraction result indicates $\mathcal{Z}^{\mathbf{U}}$ captures the rising trend of the left arm. Panel (b) illustrates individual heterogeneity across different samples. Top to bottom rows respectively show the original images $\mathbf{x}_{ij}$, reconstructed representations from $g(\mathbf{z}_{ij}^{\mathbf{V}})$, and the subtraction between adjacent reconstructions. The subtraction result showcases $\mathcal{Z}^{\mathbf{V}}$ captures variability of other limb positions.

Conversely, a larger L preserves more information during encoding, potentially improving reconstruction quality of our autoencoder framework. In summary, we need to balance the trade-off between performance of mixed-effects modeling and orthogonality, and reconstruction quality. For simulation experiments, we referenced related studies that utilized the same dataset [15], [29]. We followed their setting and adopted $L = 4$. For our clinical experiments, we considered the intricate nature of brain anatomy and thus opted for larger L. We empirically found that exceeding $L = 64$ led to a significant increase in computational time and diminished performance. Hence, we conducted a grid search over values that are powers of 2, ranging from 8 to 64,which can facilitate hardware efficiency and enhance computational performance. Ultimately, we settled on $L = 16$, as it struck an optimal balance.

The hyperparameter λ governs intensity of the orthogonality constraint on $\mathbf{U}$ and $\mathbf{V}$. As the dimensionality L increases, maintaining orthogonality between $\mathbf{U}$ and $\mathbf{V}$ becomes more challenging. Consequently, we applied a stronger constraint to ensure orthogonality. To determine the optimal value of λ, we conducted a grid search during both simulation and clinical experiments. For the simulation experiments, we explored values ranging from 0.5 to 8.0, in increments of 0.5. In contrast, for the clinical experiments, we extended the search range from 1.0 to 16.0, in increments of 1. Ultimately, we selected $\lambda = 1$ for the simulation dataset and $\lambda = 10$ for the clinical datasets. This tailored approach accommodates the specific characteristics of each dataset and ensures effective model regularization by adjusting the strength of the constraint accordingly.

In Appendix-A, we provide a comprehensive elucidation of the grid search results and the details to determine the optimal values of hyperparameters.

E. Experimental Environment

We utilize the ADAM optimizer alongside a mini-batch size of 128 and an initial learning rate of 0.001 for our experiments. *UOMETM* is implemented using Python version 3.8.16, relying on PyTorch v1.6.0 with CUDA 10.2 for GPU acceleration. The execution of *UOMETM* is performed on an NVIDIA Tesla V100SXM2 GPU with 32GB of RAM.

V. Results

In this section, we aimed to answer the following research questions in order to comprehensively evaluate our model across various aspects. **RQ1:** Does each component of *UOMETM* work as expected? **RQ2:** Can *UOMETM* outperform baseline methods on fundamental metrics that assess the performance of longitudinal trajectory modeling and orthogonality of representations? **RQ3:** Can the orthogonal representations accurately capture the global characteristics and individual heterogeneity of longitudinal progression? **RQ4:** What is the performance of *UOMETM* when it is trained in one dataset but applied to the other dataset? **RQ5:** How effectively does *UOMETM* perform on extrapolation-based downstream tasks compared to the baseline models?

The RQs mentioned above align with the subsequent subsections in order. Experiments for RQ1 were done solely on the simulation dataset due to its comprehensive nature with known ground truths. The OASIS-3 dataset was employed in RQ4, serving as an extra test set. As for other RQs, investigations were conducted on the ADNI dataset to show *UOMETM*'s clinical applicability. For brevity, we have $\mathcal{Z}$, $\mathcal{Z}^{\mathbf{U}}$, and $\mathcal{Z}^{\mathbf{V}}$ denoted as three representation spaces that include the latent representations $\mathbf{z}_{ij}$ and the orthogonal representations $\mathbf{z}_{ij}^{\mathbf{U}}$ and $\mathbf{z}_{ij}^{\mathbf{V}}$ ($\forall\, j \in [n_i], i \in [m]$), respectively. Note that all results presented here came from the test data, unless explicitly stated otherwise.

A. Validation of the Proposed Model on the Simulation Dataset

We validated the effectiveness of each component within *UOMETM* using the simulation dataset, as its known ground truths facilitate quantitative assessments.

1) $\mathcal{Z}^{\mathbf{U}}$ and $\mathcal{Z}^{\mathbf{V}}$: We visualized longitudinal features captured within the orthogonal representations from $\mathcal{Z}^{\mathbf{U}}$ and $\mathcal{Z}^{\mathbf{V}}$. To better visualize limb motion, we proposed a subtraction between adjacent images, through which we can determine whether a limb moves from negative to positive values, or whether it remains static at zero values. Based on the trained model, we reconstructed the global trajectory from $\mathcal{Z}^{\mathbf{U}}$ at various timepoints, that is, $\Psi(\mathbf{z}_{ij}^{\mathbf{U}})$, $j \in [10]$. Fig. 4(a) shows that the $\mathcal{Z}^{\mathbf{U}}$ space encodes the rising motion of the left arm while other limbs are static. Similarly, we reconstructed

TABLE II
COMPARISON OF *UOMETM* WITH THE BASELINE MODELS ON THE ADNI DATASET. NOTABLY, ONLY RESULTS IN THE FIRST ROW PERTAIN TO THE TRAINING SET.NOTE THAT * DENOTES THE STATISTICAL SIGNIFICANCE BETWEEN *UOMETM* AND A CORRESPONDING BASELINE MODEL BY TWO-SAMPLE *t*-TESTS, THAT IS, $p < \mathbf{0.05}$

Metric	β-VAE	ML-VAE	Rank-VAE	Riem-VAE	LNE	*UOMETM*
MSE of reconstruction (training) ($\times 10^{-2}$)	5.331±0.157*	4.667±0.162*	4.016±0.111*	3.686±0.127*	3.374±0.082*	**3.237±0.079**
MSE of reconstruction ($\times 10^{-2}$)	5.956±0.173*	4.991±0.157*	4.570±0.105*	3.996±0.131*	3.694±0.077*	**3.528±0.087**
Orthogonality (°)	/	79.87±2.25*	/	84.68±1.02*	/	**87.32±0.67**
Spearman correlation	0.351±0.017*	0.503±0.020*	0.933±0.010	0.897±0.012*	0.871±0.014*	**0.931±0.009**

representations from $\mathcal{Z}^{\mathbf{V}}$ across different samples, that is $\Psi(\mathbf{z}_{ij}^{\mathbf{V}})$, to exhibit individual heteorgeneity. Fig. 4(b) shows the reconstruction from $\mathcal{Z}^{\mathbf{V}}$, suggesting that the left arm remains stationary while movements are evident for the other limbs. These observations suggest that $\mathcal{Z}^{\mathbf{U}}$ captures the global trajectory of the left arm, depicting the overall movement pattern, while $\mathcal{Z}^{\mathbf{V}}$ showcases variations across different samples for all limbs except the left arm, highlighting individual heterogeneity. Remarkably, these results perfectly consistent with our expectations of the simulated data indicate that there is almost no overlap between these two spaces.

2) Cross-Reconstruction Loss: We conducted an ablation analysis on the cross-reconstruction loss (i.e., $\mathcal{L}_{\text{cross-recon}}$) to verify its effectiveness to generate the consistent global trajectory. The consistency was quantitatively assessed through Eq. (9). Under the effect of $\mathcal{L}_{\text{cross-recon}}$, the 5-fold consistency is $(2.54 \pm 0.17) \times 10^{-4}$, which indicates excellent similarity of the global trajectory among all sample pairs from $\mathcal{S}$. But after removing $\mathcal{L}_{\text{cross-recon}}$, the resultant consistency increases to $(1.30 \pm 0.35) \times 10^{-3}$, manifesting worse consistency ($p = 2.2 \times 10^{-8}$). This undesirable outcome emphasizes the significance of our introduced $\mathcal{L}_{\text{cross-recon}}$ in constructing robust orthogonal representation spaces.

3) Extrapolation From Mixed-Effects Modeling: We quantitatively assessed the extrapolation performance of our model compared to Riem-VAE. In this experiment, each subject had 10 known observations. We kept the first k and removed other observations from each subject. The missing rate, denoted as p, was calculated via $p = (10 - k)/10$. We then employed pretrained *UOMETM* and Riem-VAE to on latent representations at earlier known timepoints to compute the individual trajectories for each subject, based on which we extrapolated missing images at later timepoints using their respective time information. MSE between the extrapolated images with their ground truth was used to evaluate the accuracy of the extrapolation. In addition, we also computed MSE between the ground truth and the images computed from *UOMETM* trained based on all observations without missingness and considered it as a benchmark value.

Fig. 5 shows MSE at different missing rates. When the missing rate p increases from 10% to 70%, MSE increases for both our *UOMETM* and Riem-VAE. Nevertheless, Riem-VAE exhibits significantly higher MSE than *UOMETM* in all missing rates (all $p < 0.05$), displaying a larger deviation from the benchmark MSE value which is indicated by the dashed line in Fig. 5. This suggests better performance of *UOMETM* compared to Riem-VAE in the extrapolation task, indicating that our mixed-effects modeling functions as intended.

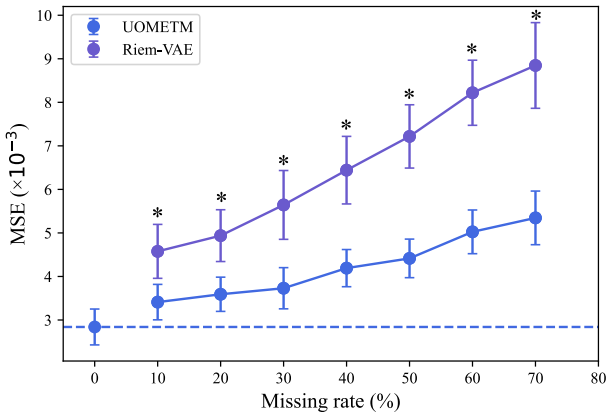


Fig. 5. MSE of extrapolation of *UOMETM* and Riem-VAE at different missing rates. The dashed line represents a benchmark value computed from *UOMETM* trained on all observations without missingness. The error bars represent standard deviations. Note that * denotes the statistically significant difference between two models by two-sample *t*-tests, that is, $p < 0.05$.

B. Validation of UOMETM on the ADNI Dataset

To demonstrate the superiority of *UOMETM*, we assessed its performance alongside the baseline models on the ADNI dataset using three fundamental metrics. First, we assessed the quality of reconstruction using MSE. Second, we quantified the orthogonality between $\mathcal{Z}^{\mathbf{U}}$ and $\mathcal{Z}^{\mathbf{V}}$ by calculating the angle of the representations from these two spaces. Third, we computed the monotonicity of the global trajectory in relation with age using the Spearman correlation.

Table II lists three metrics. The MSEs of reconstruction for training and test datasets are $(3.237 \pm 0.079) \times 10^{-2}$ and $(3.528 \pm 0.087) \times 10^{-2}$, which are significantly lower than any other baseline model (all $p < 0.05$).

The angle between *UOMETM*'s representation spaces is $87.32° \pm 0.67°$, significantly surpassing both Riem-VAE ($p = 2.1 \times 10^{-6}$) and ML-VAE ($p = 8.5 \times 10^{-9}$). The proximity of *UOMETM*'s angle to 90° signifies excellent orthogonality, suggesting better separation between the global trajectory space and the space of individual trajectories.

Furthermore, we utilized the Spearman correlation to examine the monotonic relationship between clinical age t_{ij} and representations $\mathbf{z}_{ij}^{\mathbf{U}}$. Here, principal component analysis (PCA) was performed on $\mathbf{z}_{ij}^{\mathbf{U}}$ and its first PC was used to correlate

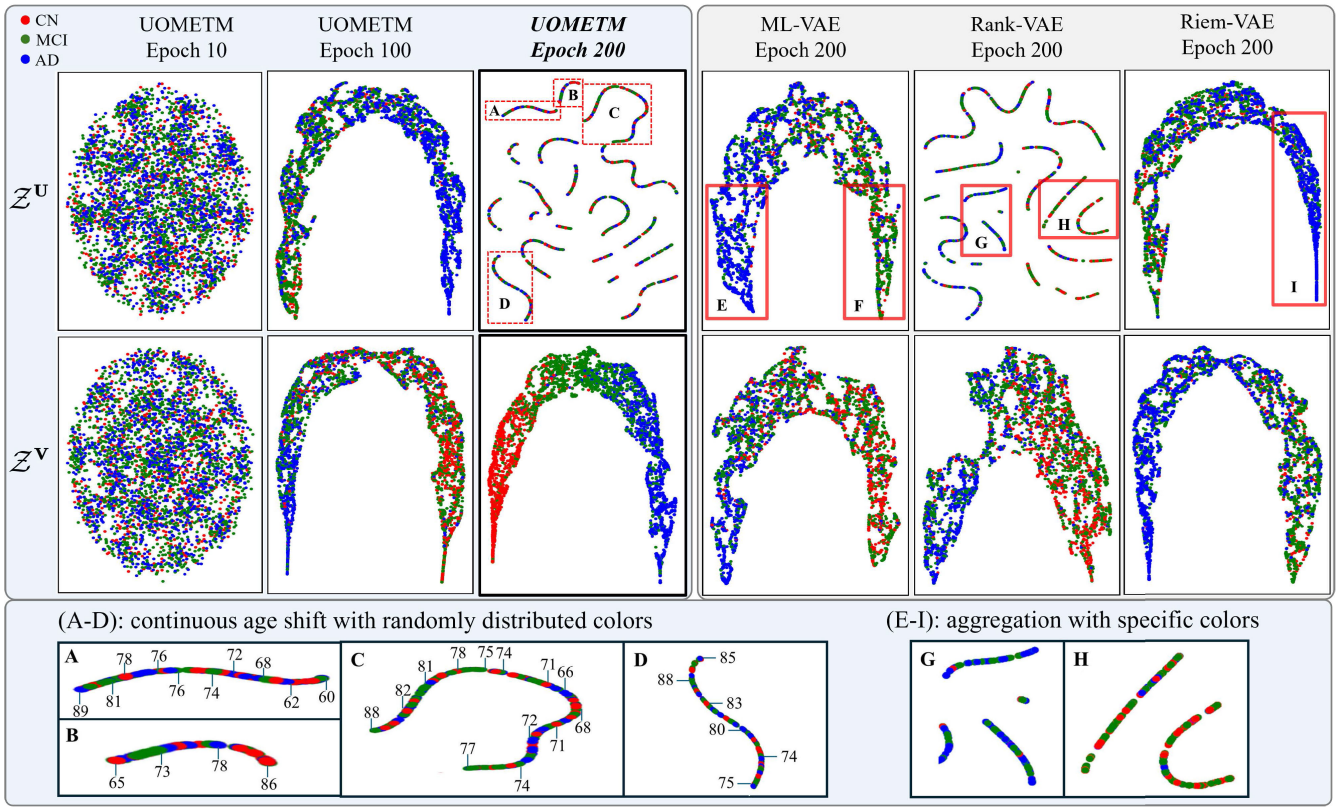


Fig. 6. Visualization of feature distribution in representation spaces $\mathcal{Z}^{\mathbf{U}}$ and $\mathcal{Z}^{\mathbf{V}}$. The left panel traces the evolution over epochs, suggesting progressive learning of global patterns across all diagnostic groups in $\mathcal{Z}^{\mathbf{U}}$ and individual heterogeneity related AD diagnosis in $\mathcal{Z}^{\mathbf{V}}$. The zoomed-in views of the red dashed rectangles (A-D) show that subjects in $\mathcal{Z}^{\mathbf{U}}$ are clustered, with age shifting continuously and diagnostic labels randomly distributed. The right panel displays the final outputs of the baseline models. Red rectangles (E-I) highlight the aggregation of specific diagnostic groups, contradicting to our expectations.

with clinical age. *UOMETM* demonstrates correlations of 0.931 ± 0.009, significantly excelling all other baseline models (all $p < 0.05$), except Rank-VAE ($p = 0.64$). The higher correlations of *UOMETM* and Rank-VAE indicate more accurate representations of the global trajectory and more precise longitudinal progressions captured by them. Notably, Rank-VAE achieves its high performance by incorporating a tailored ranking loss to enforce the chronological consistency of the global trajectory. Our model, while presenting an approximate accuracy, showcases exceptional ability in capturing the inherent underlying global progression order within the data.

In summary, *UOMETM* consistently outperforms baseline models across key metrics, quantitatively affirming the superior ability of *UOMETM* to capture and separate longitudinal global and individual patterns.

C. Orthogonal Representations

We comprehensively assessed the orthogonal representations to demonstrate their effectiveness in terms of characterizing global and individual trajectories.

First, we employed a manifold discovery and analysis (MDA) method [34] on the representation spaces $\mathcal{Z}^{\mathbf{U}}$ and $\mathcal{Z}^{\mathbf{V}}$, allowing us to visualize the distribution within them while maintaining the local geometry of their space manifolds. This technique can reveal the appropriateness, generalizability, and adversarial robustness of our deep neural network. The results in Fig. 6 demonstrate the evolution of latent representations from the disordered states to the ordered ones presenting clear patterns across epochs. When the training just starts, representations are randomly distributed in both spaces. Over increasing epochs, these representations tend to organize themselves progressively, yet with a few differences: in $\mathcal{Z}^{\mathbf{U}}$, they intermingle in color; contrastingly, in $\mathcal{Z}^{\mathbf{V}}$, a continuous color shift occurs. Eventually, representations in $\mathcal{Z}^{\mathbf{U}}$ from three diagnostic groups aggregate with age shifting continuously and disseminate randomly within each cluster, while those in $\mathcal{Z}^{\mathbf{V}}$ tend to illustrate a more perceptible shift in color with obvious aggregation of the same diagnostic groups. These results suggest a learning progress of global patterns across all diagnostic groups in $\mathcal{Z}^{\mathbf{U}}$ and individual heterogeneity related AD diagnosis in $\mathcal{Z}^{\mathbf{V}}$. Additionally, we showcase the final outputs of the baseline models. Their spaces of $\mathcal{Z}^{\mathbf{U}}$ demonstrate a distinct aggregation of specific diagnostic groups, as highlighted by the red rectangles in Fig. 6, which is contradictory to our expectations. Meanwhile, the ability to capture individual heterogeneity of their spaces of $\mathcal{Z}^{\mathbf{V}}$ is obviously inferior to our model.

Second, we illustrated the space of $\mathcal{Z}^{\mathbf{U}}$ to present the consistent global trajectory across these three groups (CN, MCI, and AD). Here, we computed the global trajectories based on our model trained using all diagnostic subjects.

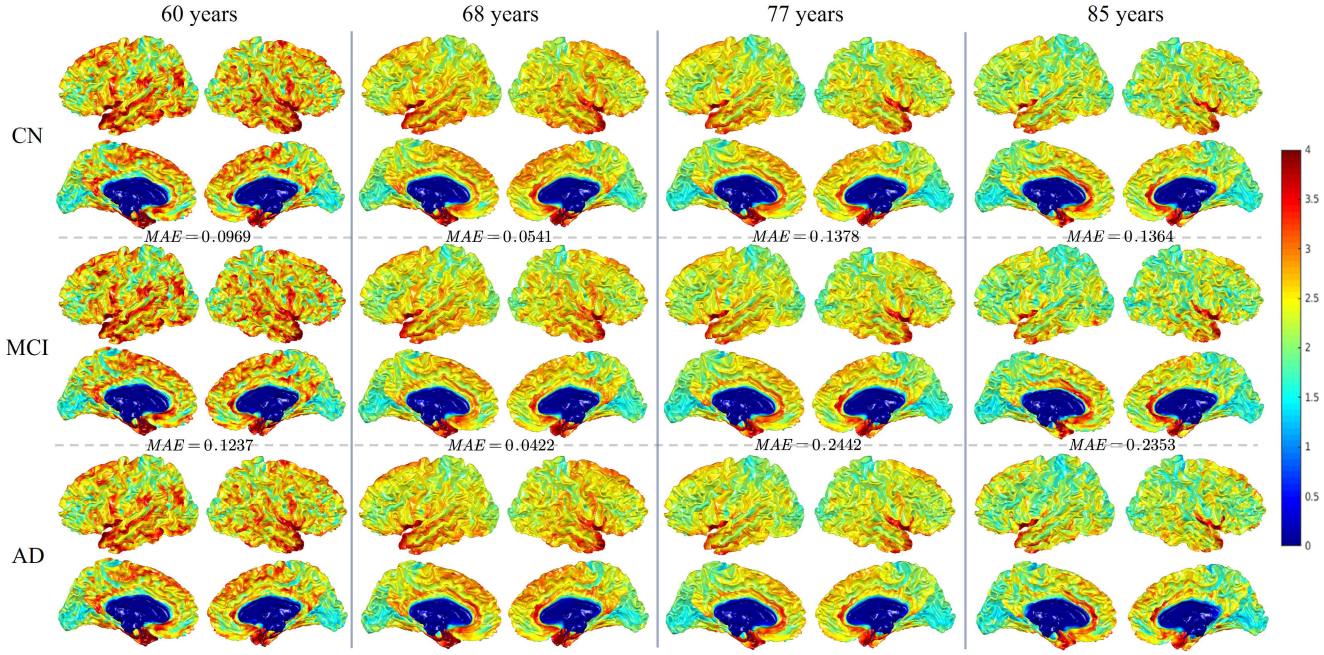


Fig. 7. Visualization of cortical thickness captured in the space $\mathcal{Z}^{\mathbf{U}}$ on a cortical surface. Each column displays results with different diagnostic labels but at the same clinical age, while each row represents the global trajectory of a specific diagnostic group. Remarkably similar results across columns highlight the successful capture of global characteristics. Note that MAE along the dashed line denotes the mean absolute error of cortical thickness sampled on both sides of the dashed line.

We expected that the global trajectories of cortical thickness were similar across all the subjects, regardless their diagnosis. Fig. 7 illustrates cortical thickness of normal controls (CN), MCI, and AD patients at different ages. Subjects with different diagnostic labels but in the same clinical age exhibit a highly similar pattern of cortical thickness. As age increases, the subjects from different groups demonstrate a comparably declining trend in cortical thickness. This suggests that the space $\mathcal{Z}^{\mathbf{U}}$ effectively excluded individual heterogeneity while manifesting a global decreasing trajectory in cortical thickness with age.

Third, we evaluated the space of $\mathcal{Z}^{\mathbf{V}}$ to demonstrate individual heterogeneity. For this purpose, we conducted an AD vs. CN classification task based on representations $\mathbf{z}_{ij}^{\mathbf{V}}$, referred to as *UOMETM-*$\mathcal{Z}^{\mathbf{V}}$, as well as the baseline models. More specifically, by leveraging our model that is pre-trained on all subjects from the training set and kept frozen, we trained a binary classifier on the fixed representations $\mathbf{z}_{ij}^{\mathbf{V}}$ from CN and AD subjects in the training set. Subsequently, we evaluated our classifier on CN and AD subjects in the test set. The classifier was designed as a multi-layer perceptron containing two FullyConnected layers and was optimized by minimizing a cross-entropy loss of the diagnostic label. To account for imbalanced samples in the two diagnostic groups, besides classification accuracy (ACC), we also computed balanced accuracy (BACC) [35] and F1 score [36] as quantitative metrics. Based on the classification results presented in Table III(a), *UOMETM-*$\mathcal{Z}^{\mathbf{V}}$ achieves satisfactory values for the specified metrics: 89.6±1.43%, 88.5±2.76%, and 88.7±2.44%. In comparison to the baseline models, *UOMETM-*$\mathcal{Z}^{\mathbf{V}}$ significantly surpasses most of them (all $p < 0.05$) except LNE ($0.5 < p < 0.1$), which underscores the effectiveness of the space $\mathcal{Z}^{\mathbf{V}}$ in preserving and harnessing essential subject-specific information for classification.

TABLE III
RESULTS OF THE AD VS. CN CLASSIFICATION TASK ON THE ADNI DATASET (a) AND THE OASIS-3 DATASET (b).NOTE THAT * DENOTES THE STATISTICAL SIGNIFICANCE BETWEEN *UOMETM-*$\mathcal{Z}^{V}$ AND A CORRESPONDING MODEL BY TWO-SAMPLE *t*-TESTS, THAT IS $p < 0.05$

(a)

Methods	ACC (%)	BACC (%)	F1 score (%)
β-VAE	81.6±3.22*	79.1±4.55*	77.3±3.75*
ML-VAE	82.0±3.18*	80.5±3.66*	78.9±3.13*
Rank-VAE	84.4±2.72*	84.8±3.73*	81.0±2.76*
Riem-VAE	86.5±2.17*	85.4±2.77*	81.9±3.03*
LNE	88.1±1.85	86.1±2.91	83.7±2.39
UOMETM-$\mathcal{Z}^{\mathbf{U}}$	63.4±4.62*	60.8±4.25*	59.2±3.96*
UOMETM-$\mathcal{Z}^{\mathbf{V}}$	**89.6±1.43**	**88.5±2.76**	**85.7±2.56**
UOMETM-$\mathcal{Z}^{\mathbf{U+V}}$	89.0±1.90	87.5±3.15	85.2±2.82

(b)

Methods	ACC (%)	BACC (%)	F1 score (%)
β-VAE	74.1±3.31*	70.3±3.43*	56.0±3.59*
ML-VAE	76.1±3.14*	72.3±3.21*	58.7±2.98*
Rank-VAE	77.8±2.61*	74.8±2.93*	62.0±2.67*
Riem-VAE	78.9±2.61*	75.6±2.75*	63.2±2.84*
LNE	80.1±2.80	77.3±2.96	65.4±2.72
UOMETM-$\mathcal{Z}^{\mathbf{U}}$	56.7±4.42*	53.9±4.55*	36.7±4.17*
UOMETM-$\mathcal{Z}^{\mathbf{V}}$	**81.5±2.82**	**78.7±2.41**	**67.6±2.84**
UOMETM-$\mathcal{Z}^{\mathbf{U+V}}$	80.9±2.79	77.7±2.52	66.2±2.84

Fourth, we assessed the leakage of individual characteristics into $\mathcal{Z}^{\mathbf{U}}$. Specifically, we performed the same AD vs. CN classification task but using representations $\mathbf{z}_{ij}^{\mathbf{U}}$ and

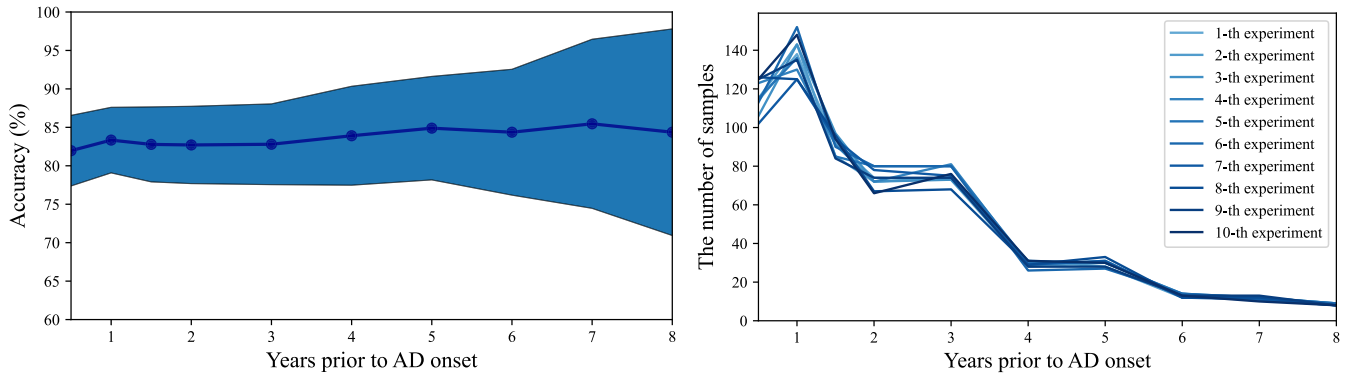


Fig. 8. Results of AD conversion forecasting tasks. Note that the forecasted timepoint is the AD onset, where conversion subjects convert from CN or MCI to AD. (a) Mean values of prediction accuracies between forecasted diagnostic labels and ground truths at different years prior to the AD onset; The shading area represents the standard deviation for the accuracy. (b) The number of conversion samples recorded at different years prior to the AD onset.

combined representations $\mathbf{z}_{ij}^{\mathbf{U}} + \mathbf{z}_{ij}^{\mathbf{V}}$, denoted as *UOMETM*-$\mathcal{Z}^{\mathbf{U}}$ and *UOMETM*-$\mathcal{Z}^{\mathbf{U+V}}$ respectively. On the one hand, *UOMETM*-$\mathcal{Z}^{\mathbf{U}}$ exhibits poor performance. On the other hand, *UOMETM*-$\mathcal{Z}^{\mathbf{U+V}}$ demonstrates comparable results to *UOMETM*-$\mathcal{Z}^{\mathbf{V}}$ in terms of all metrics (all $p > 0.1$, see Table III(a)). These two experiments both suggest that $\mathbf{z}_{ij}^{\mathbf{U}}$ only contributes noise for diagnostic features, lacking discriminative power for the classification. This observation further indicates that the two spaces are well separate, with no leakage of individual information into $\mathcal{Z}^{\mathbf{U}}$.

D. Generalization and Robustness

Our objective was to assess the generalization and robustness of *UOMETM*. To achieve this, we leveraged *UOMETM* as well as the classifier which were previously trained on the training set of the ADNI dataset. Specifically, we kept them fixed and tested them directly on the OASIS-3 dataset to classify CN and AD subjects. Note that these subjects from the OASIS-3 dataset only functioned as an extra test set, with no training process involving them. The classification results, detailed in Table III(b), show a certain decrease in accuracy compared to those obtained on the ADNI dataset. However, the model continues to perform at a satisfactory level, though OASIS-3 is an independent dataset. Furthermore, we extended this analysis to encompass other baseline models, testing them under the same conditions on the OASIS-3 dataset. It is observed that although these models also experience a drop in accuracy when transitioning from the ADNI to the OASIS-3 dataset, their relative performance hierarchy, such as performance ranking, remains consistent with that for the ADNI dataset. These observations underscore the robustness of our proposed model, demonstrating its resilience and adaptability in recognizing patterns across different datasets.

E. AD Conversion Forecasting

Forecasting the timepoint at which CN or MCI patients convert to AD is of practical importance. Therefore, we undertook a task encompassing all conversion subjects with a minimum of four timepoints from the ADNI dataset, with the objective of predicting the diagnostic label at the onset of Alzheimer's disease (AD). This aimed to rigorously evaluate the extrapolation capability of our longitudinal model.

The forecasting task can be articulated mathematically as follows: for the i-th subject, we forecasted the diagnostic label at the AD onset, denoted as the l-th visit, given k_i known data $\{\mathbf{x}_{ij} \mid j \in [k_i]\}$ and age information at the target timepoint t_{il}. The year prior to the AD onset associated with this forecasting task was $(t_{il} - t_{ik_i})$. It is worth noting that a single subject could generate multiple conversion samples with different years prior to AD onset as the quantity of available data k_i could vary under the constraint $4 \leq k_i < l \leq n_i$. Before implementing the forecasting task, we trained *UOMETM* on all subjects in the training set from the ADNI dataset, subsequently fixing all the model parameters. Then for every eligible subject i in the training set, we utilized the frozen parameters of *OMETM* to compute the individual trajectory from representations at earlier k_i timepoints. Through this trajectory, we extrapolated the representation at the AD onset (that is, $\mathbf{z}_{il}$) and fed it into a FullyConnected layer to predict its diagnostic label as AD or CN. The optimization of the FullyConnected layer was achieved through minimizing the cross-entropy loss of diagnostic labels. The forecasting performance was evaluated on eligible subjects in the test set through the prediction accuracy between forecasted diagnostic labels and ground truths.

We endeavored to illustrate the efficacy of our model in forecasting AD onset across varying time intervals prior to the onset event, referred to as "years prior to the AD onset" in our study. To achieve this, we visualized the prediction accuracy of the AD conversion as a function of years prior to the AD onset. The results in Fig. 8(a) illustrate the mean values of accuracies with a shading area representing standard deviations at different years prior to the AD onset. The mean accuracy of the AD conversion is relatively consistent, exhibiting a range of 81.96% to 85.46%. The standard deviations increase over time due to the drop in the number of subjects available (see Fig. 8(b)). Across all subjects, *UOMETM* achieve accuracies of 83.02±2.89%, significantly surpassing 79.24±2.62%

TABLE IV
GRID SEARCH ON λ OF THE SIMULATION EXPERIMENTS

λ ($L = 4$)	$\mathcal{L}_{\text{model}}$	$\mathcal{L}_{\text{ortho}}$	$\mathcal{L}_{\text{model}} + \frac{\lambda}{2}\mathcal{L}_{\text{ortho}}$
0.5	0.161±0.018	0.090±0.013	0.184±0.026
1.0	0.161±0.020	0.044±0.019	**0.183±0.026**
1.5	0.162±0.022	0.040±0.010	0.192±0.026
2.0	0.163±0.021	0.040±0.012	0.203±0.027
2.5	0.166±0.021	0.039±0.010	0.215±0.028
3.0	0.170±0.020	0.038±0.009	0.227±0.026
3.5	0.175±0.022	0.038±0.009	0.224±0.029
4.0	0.177±0.024	0.037±0.011	0.251±0.030
4.5	0.181±0.023	0.036±0.010	0.262±0.028
5.0	0.186±0.024	0.035±0.010	0.274±0.029
5.5	0.189±0.025	0.034±0.008	0.283±0.027
6.0	0.195±0.027	0.032±0.008	0.291±0.027
6.5	0.197±0.026	0.032±0.009	0.301±0.029
7.0	0.202±0.027	0.030±0.010	0.307±0.030
7.5	0.205±0.028	0.030±0.008	0.318±0.030
8.0	0.206±0.028	0.028±0.009	0.318±0.033

($p = 0.007$) obtained from Riem-VAE. These results highlight the superiority of our *UOMETM* in flexibly capturing the longitudinal progression patterns, leading to an enhanced forecasting performance in the AD diagnostic task.

To sum up, *UOMETM* is endowed with excellent extrapolation capacity, which enables us to effectively handle missing data and estimate future trends with several known data points, making our model highly versatile and valuable in clinical longitudinal tasks.

VI. CONCLUSION

In this study, we propose *Unsupervised Orthogonal Mixed-Effects Trajectory Modeling (UOMETM)*. We successfully model a global trajectory and individual trajectories for longitudinal data through mixed-effects modeling, and we enrich these trajectories with high-level interpretability via innovative orthogonal representations. The performance of our proposed model is comprehensively evaluated using both a simulation dataset and two clinical datasets, with in-depth assessments of the orthogonal representations and longitudinal properties compared to the SOTA baseline methods. The experiments on the simulation dataset affirm the successful validation of each component within *UOMETM*, with all elements functioning as intended. The empirical results on the clinical datasets demonstrate that the effectiveness of the orthogonality leads to higher and more robust accuracy in AD classification. Furthermore, our results underscore the remarkable extrapolation ability through AD diagnostic forecasting tasks, further enriching the model's capabilities. In summary, *UOMETM* shows promising potential for capturing longitudinal progression and provides valuable insights into disease progression modeling and supporting clinical decision-making.

APPENDIX

A. Hyperparameter Selection

In this section, we will provide comprehensive descriptions for determining the values of hyperparameters, namely L and λ, separately for both simulation experiments and clinical experiments.

TABLE V
GRID SEARCH ON L AND λ OF THE CLINICAL EXPERIMENTS. (A) RECONSTRUCTION QUALITY: $L_{\text{SELF-}recon}(\times 10^{-2})$. (B) PERFORMANCE OFMIXED-EFFECTS MODELING AND ORTHOGONALITY: $L_{\text{MODEL}} + \frac{\lambda}{2}L_{\text{ORTHO}}$. NOTE THAT *DENOTES THE STATISTICAL SIGNIFICANCE BETWEEN ALL RESULTS OF $L = k$ AND $L = 2k$ BY TWO-SAMPLE t-TESTS, THAT IS $p < 0.05$ ACROSS ALL λ VALUES

(a)

	$L = 8$*	$L = 16$*	$L = 32$*	$L = 64$
$\lambda = 1$	4.321±0.095	3.465±0.093	3.282±0.079	3.148±0.073
$\lambda = 2$	4.327±0.092	3.470±0.095	3.290±0.083	3.155±0.068
$\lambda = 3$	4.335±0.101	3.479±0.078	3.295±0.092	3.160±0.079
$\lambda = 4$	4.342±0.089	3.485±0.089	3.302±0.085	3.164±0.081
$\lambda = 5$	4.351±0.102	3.492±0.097	3.308±0.085	3.171±0.064
$\lambda = 6$	4.358±0.103	3.501±0.086	3.313±0.089	3.179±0.078
$\lambda = 7$	4.365±0.095	3.509±0.096	3.319±0.078	3.185±0.087
$\lambda = 8$	4.373±0.098	3.514±0.076	3.325±0.084	3.192±0.081
$\lambda = 9$	4.379±0.100	3.520±0.088	3.331±0.090	3.201±0.064
$\lambda = 10$	4.387±0.092	3.528±0.087	3.338±0.084	3.207±0.063
$\lambda = 11$	4.399±0.095	3.536±0.092	3.344±0.081	3.215±0.068
$\lambda = 12$	4.415±0.091	3.551±0.095	3.350±0.094	3.227±0.073
$\lambda = 13$	4.431±0.098	3.563±0.085	3.360±0.088	3.238±0.078
$\lambda = 14$	4.447±0.107	3.580±0.088	3.373±0.089	3.250±0.082
$\lambda = 15$	4.463±0.106	3.593±0.093	3.385±0.090	3.259±0.076
$\lambda = 16$	4.480±0.100	3.609±0.093	3.399±0.085	3.263±0.089

(b)

	$L = 8$*	$\boldsymbol{L = 16}$*	$L = 32$*	$L = 64$
$\lambda = 1$	0.767±0.054	1.296±0.087	5.803±0.145	14.212±0.632
$\lambda = 3$	0.750±0.054	1.247±0.080	5.710±0.129	13.997±0.412
$\lambda = 4$	0.736±0.058	1.203±0.089	5.652±0.152	13.702±0.465
$\lambda = 5$	0.727±0.053	1.167±0.097	5.629±0.118	13.561±0.524
$\lambda = 6$	0.722±0.049	1.132±0.086	5.594±0.164	13.473±0.367
$\lambda = 7$	0.723±0.060	1.109±0.096	5.569±0.137	13.424±0.541
$\lambda = 8$	0.726±0.059	1.088±0.081	5.545±0.127	13.391±0.467
$\lambda = 9$	0.732±0.054	1.078±0.088	5.530±0.128	13.357±0.617
$\lambda = 10$	0.737±0.056	**1.072±0.087**	5.524±0.119	13.302±0.581
$\lambda = 11$	0.747±0.053	1.075±0.092	5.519±0.150	13.275±0.637
$\lambda = 12$	0.759±0.061	1.086±0.095	5.516±0.174	13.258±0.527
$\lambda = 13$	0.782±0.062	1.100±0.085	5.510±0.134	13.243±0.682
$\lambda = 14$	0.799±0.058	1.113±0.088	5.521±0.124	13.228±0.656
$\lambda = 15$	0.825±0.058	1.148±0.093	5.535±0.164	13.223±0.727
$\lambda = 16$	0.860±0.060	1.201±0.093	5.547±0.173	13.219±0.627

1) Simulation Experiments: In sectionIV-D, we mentioned that we fixed $L = 4$ to align with the settings of previous studies utilizing the same dataset. Consequently, only the hyperparameter λ remains to be determined. To address this, we conducted a grid search across values from 0.5 to 8.0, with increments of 0.5. As λ governed the strength of $\mathcal{L}_{\text{ortho}}$ within Eq. (11), and our optimization objective focused on minimizing Eq. (11), we employed the value of Eq. (11) as the metric for evaluating performance across different λ values.

The results are illustrated in Table IV. We notice that as λ increases, the values of $\mathcal{L}_{\text{model}}$ tend to rise while those of $\mathcal{L}_{\text{ortho}}$ decrease, consistent with our expectations. The evaluation metric results exhibit a U-shaped pattern, with the lowest value achieved at $\lambda = 1$. Consequently, we set $\lambda = 1$ accordingly.

2) Clinical Experiments: We utilized grid search to determine suitable values for L and λ, considering $L = \{8, 16, 32, 64\}$ and $\lambda = \{1, 2, \ldots, 16\}$, as outlined in sectionIV-D. As discussed earlier, increasing L enhances reconstruction quality but may compromise mixed-effects modeling and orthogonality performance. Hence, L should be determined such that

a balance between these factors must be struck, considering both $\mathcal{L}_{\text{self-recon}}$ and Eq. (11). Similarly to the simulation experiments, we aimed to minimize Eq. (11) to determine λ.

The reconstruction quality outcomes under different L and λ values are presented in Table V(a). Doubling L leads to a significant enhancement in reconstruction quality with a notable decrease in reconstruction errors ($p < 0.05$ across all λ values). However, $L = 8$ shows an apparent inferior performance compared to $L = \{16, 32, 64\}$, which exhibit closer satisfactory results despite significant differences. The impact of λ is evident, where smaller values consistently yield superior reconstruction quality. However, the impact of varying λ on performance intensity remains modest. Consequently, we exclude $L = 8$ from further consideration and focus on alternatives from $\{16, 32, 64\}$.

The performance of mixed-effects modeling and orthogonality under various L and λ settings is depicted in Table V(b). We note that doubling L results in a significant decrease in performance with a marked increase in Eq. (11) values ($p < 0.05$ across all λ values). Notably, $L = \{32, 64\}$ exhibit inferior performance compared to $L = \{8, 16\}$, which demonstrate acceptable results. Regarding the influence of λ given a fixed L, we observe a U-shaped pattern akin to the simulation experiments. However, changes in λ have a limited impact on performance intensity. Thus, we eliminate $L = \{32, 64\}$ from consideration and concentrate on alternatives from $\{8, 16\}$.

Considering these findings, we set $L = 16$. Given this value, we select λ to minimize Eq. (11), resulting in $\lambda = 10$.

REFERENCES

[1] E. J. Caruana, M. Roman, J. Hernández-Sánchez, and P. Solli, "Longitudinal studies," *J. Thoracic Disease*, vol. 7, no. 11, p. E537, Nov. 2015.

[2] K. L. Mills and C. K. Tamnes, "Methods and considerations for longitudinal structural brain imaging analysis across development," *Develop. Cognit. Neurosci.*, vol. 9, pp. 172–190, Jul. 2014.

[3] I. Sintini et al., "Longitudinal neuroimaging biomarkers differ across Alzheimer's disease phenotypes," *Brain*, vol. 143, no. 7, pp. 2281–2294, Jul. 2020.

[4] A. R. Roeckner, K. I. Oliver, L. A. M. Lebois, S. J. H. van Rooij, and J. S. Stevens, "Neural contributors to trauma resilience: A review of longitudinal neuroimaging studies," *Transl. Psychiatry*, vol. 11, no. 1, p. 508, Oct. 2021.

[5] E. Girden, *ANOVA*. Newbury Park, CA, USA: SAGE, Nov. 1992.

[6] P. Schober and T. R. Vetter, "Repeated measures designs and analysis of longitudinal data: If at first you do not succeed—Try, try again," *Anesthesia Analgesia*, vol. 127, no. 2, pp. 569–575, Aug. 2018.

[7] L. Frings, I. Mader, B. G. Landwehrmeyer, C. Weiller, M. Hüll, and H. Huppertz, "Quantifying change in individual subjects affected by frontotemporal lobar degeneration using automated longitudinal MRI volumetry," *Hum. Brain Mapping*, vol. 33, no. 7, pp. 1526–1535, Jul. 2012.

[8] J. L. Bernal-Rusiel, D. N. Greve, M. Reuter, B. Fischl, and M. R. Sabuncu, "Statistical analysis of longitudinal NeuroImage data with linear mixed effects models," *NeuroImage*, vol. 66, pp. 249–260, Feb. 2013.

[9] A. Bône, O. Colliot, and S. Durrleman, "Learning distributions of shape trajectories from longitudinal datasets: A hierarchical model on a manifold of diffeomorphisms," 2018, *arXiv:1803.10119*.

[10] J.-B. Schiratti, S. Allassonniere, O. Colliot, and S. Durrleman, "Learning spatiotemporal trajectories from manifold-valued longitudinal data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–11.

[11] M. Louis, R. Couronné, I. Koval, B. Charlier, and S. Durrleman, "Riemannian geometry learning for disease progression modelling," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Apr. 2019, pp. 542–553.

[12] S. Gruffaz, P.-E. Poulet, E. Maheux, B. Jedynak, and S. Durrleman, "Learning Riemannian metric for disease progression modeling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 23780–23792.

[13] J. Du, A. Goh, S. Kushnarev, and A. Qiu, "Geodesic regression on orientation distribution functions with its application to an aging study," *NeuroImage*, vol. 87, pp. 416–426, Feb. 2014.

[14] A. Kolesnikov, X. Zhai, and L. Beyer, "Revisiting self-supervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1920–1929.

[15] R. Couronné, P. Vernhet, and S. Durrleman, "Longitudinal self-supervision to disentangle inter-patient variability from disease progression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., Cham, Switzerland: Springer, 2021, pp. 231–241.

[16] Q. Zhao, Z. Liu, E. Adeli, and K. M. Pohl, "Longitudinal self-supervised learning," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102051.

[17] J. Ouyang, Q. Zhao, E. Adeli, G. Zaharchuk, and K. M. Pohl, "Self-supervised learning of neighborhood embedding for longitudinal MRI," *Med. Image Anal.*, vol. 82, Nov. 2022, Art. no. 102571.

[18] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[19] Y. Liu, Z. Dong, P. Zhu, and S. Liu, "Unsupervised underwater image enhancement based on feature disentanglement," *J. Electron. Inf. Technol.*, vol. 44, no. 10, pp. 3389–3398, 2022.

[20] S. Liu, K.-H. Thung, L. Qu, W. Lin, D. Shen, and P.-T. Yap, "Learning MRI artefact removal with unpaired data," *Nature Mach. Intell.*, vol. 3, no. 1, pp. 60–67, Jan. 2021.

[21] P. Zhu, Y. Liu, Y. Wen, M. Xu, X. Fu, and S. Liu, "Unsupervised underwater image enhancement via content-style representation disentanglement," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106866.

[22] R. Wang, Q. Zhou, and G. Zheng, "EDRL: Entropy-guided disentangled representation learning for unsupervised domain adaptation in semantic segmentation," *Comput. Methods Programs Biomed.*, vol. 240, Oct. 2023, Art. no. 107729.

[23] P. Zhu, Y. Liu, M. Xu, X. Fu, N. Wang, and S. Liu, "Unsupervised multiple representation disentanglement framework for improved underwater visual perception," *IEEE J. Ocean. Eng.*, vol. 49, no. 1, pp. 48–65, Jan. 2024.

[24] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," 2018, *arXiv:1812.05069*.

[25] I. Higgins et al., "β-VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. ICLR*, 2017. [Online]. Available: https://openreview.net/forum?id=Sy2fzU9gl

[26] D. Bouchacourt, R. Tomioka, and S. Nowozin, "Multi-level variational autoencoder: Learning disentangled representations from grouped observations," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Apr. 2018, pp. 2095–2102.

[27] I. Koval et al., "AD course map charts Alzheimer's disease progression," *Sci. Rep.*, vol. 11, no. 1, p. 8020, Apr. 2021.

[28] B. Sauty and S. Durrleman, "Riemannian metric learning for progression modeling of longitudinal datasets," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5.

[29] B. Sauty and S. Durrleman, "Progression models for imaging data with longitudinal variational auto encoders," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*. Cham, Switzerland: Springer, 2022, pp. 3–13.

[30] I. Higgins et al., "Towards a definition of disentangled representations," 2018, *arXiv:1812.02230*.

[31] F. N. Gumedze and T. T. Dunne, "Parameter estimation and inference in the linear mixed model," *Linear Algebra Appl.*, vol. 435, no. 8, pp. 1920–1944, Oct. 2011.

[32] J. Du, L. Younes, and A. Qiu, "Whole brain diffeomorphic metric mapping via integration of Sulcal and Gyral curves, cortical surfaces, and images," *NeuroImage*, vol. 56, no. 1, pp. 162–173, May 2011.

[33] C. Liu, H. Ji, and A. Qiu, "Fast vertex-based graph convolutional neural network and its application to brain images," *Neurocomputing*, vol. 434, pp. 1–10, Apr. 2021.

[34] M. T. Islam et al., "Revealing hidden patterns in deep neural network feature space continuum via manifold learning," *Nature Commun.*, vol. 14, no. 1, p. 8506, Dec. 2023.

[35] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3121–3124.

[36] Y. Sasaki, "The truth of the f-measure," *Teach Tutor Mater*, vol. 1, no. 5, pp. 1–5, Jan. 2007.