

A Implementation details

This section provides an overview of the design choices and implementation details for BrainODE. We begin by detailing the longitudinal brain shape reconstruction process from 3D brain MRI scans, which involves segmentation and reconstruction of the hippocampal and lateral ventricle (LV) shapes. Subsequently, we evaluate the effectiveness of principal component analysis for representing variable brain shapes, reducing dimensionality while preserving key shape features. Afterward, we present details of the neural architecture of BrainODE and its training strategies. Our source code, including data preprocessing scripts and model configurations, is publicly available at <https://anonymous.4open.science/r/BrainODE-7558>

A.1 Longitudinal brain shape reconstruction

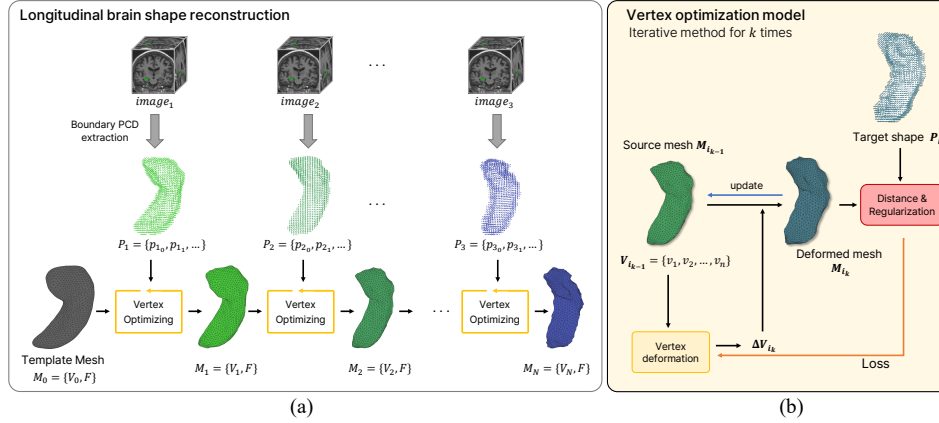


Figure 4: Illustration of brain shape reconstructions for a single subject by iteratively deforming a template mesh to match longitudinal MRI scans.

We reconstruct the longitudinal brain shapes $V = \{V_1, V_2, \dots, V_n\}$ by modifying the brain shapes modeling method proposed by Park *et al.* [11]. Their approach reconstructs cross-sectional individual brain shapes by iteratively deforming a template mesh to align with target shapes, which are represented as point clouds extracted from the boundaries of segmented 3D brain MRI masks. To adapt this cross-sectional method for longitudinal analysis, we sequentially update a template mesh M_0 over time, using the deformed mesh from the previous time point, M_{i-1} , as the initial mesh for the next step, M_i . The process begins with the first observed MRI scan and proceeds iteratively through subsequent time points, optimizing vertex positions to minimize both distance and regularization loss, as illustrated in Figure 4.

As part of this pipeline, we use SynthSeg [1] to segment brain regions from MRI scans. From the boundaries of the segmented masks, point clouds P_i representing the target shapes are constructed. The initial shape V_1 is obtained by deforming the template mesh vertices V_0 to fit the target point cloud P_1 . Then, each subsequent shape V_i is reconstructed by deforming the previous mesh V_{i-1} to align with the corresponding target P_i , using a vertex optimization model.

The vertex optimization model (Figure 4(b)) follows the optimization strategy of Point2Mesh [4], which iteratively updates the vertices to align with the target point cloud. The objective function \mathcal{L} consists of two components: 1) a distance loss \mathcal{L}_{dist} , which measures the discrepancy between the deformed mesh V_i and the target point cloud P_i , and 2) a regularization loss \mathcal{L}_{reg} , which encourages smooth and plausible deformations.

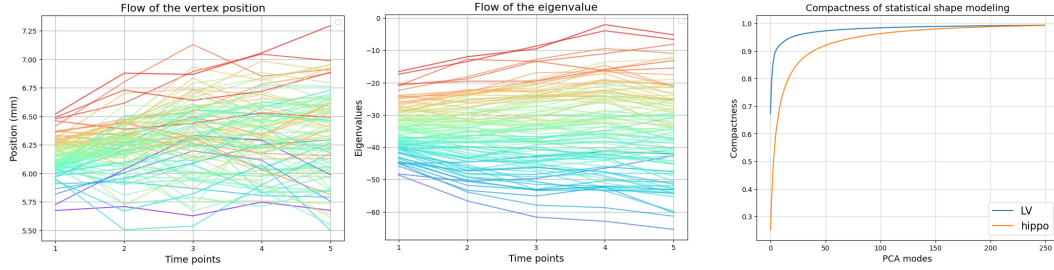
Specifically, the distance loss \mathcal{L}_{dist} combines the Chamfer distance \mathcal{L}_{cf} between V_i and P_i , and the point-to-face distance \mathcal{L}_{pm} between P_i and the mesh surface M_i (i.e., $\mathcal{L}_{dist} = \lambda_{cf}\mathcal{L}_{cf} + \lambda_{pm}\mathcal{L}_{pm}$). The regularization loss \mathcal{L}_{reg} includes vertex displacement $\|\Delta_{vert}\|_2$, normal displacement $\|\Delta_{norm}\|_2$, edge length variance \mathcal{L}_{edge} to prevent mesh distortion, normal consistency $\mathcal{L}_{cons}(norm)$, and Laplacian loss \mathcal{L}_{lap} for surface smoothness. The final form of the loss function used in the vertex optimization model is defined as follows:

$$\mathcal{L} = \mathcal{L}_{dist} + \mathcal{L}_{reg} \quad (5)$$

where $\mathcal{L}_{reg} = \lambda_{vert} \|\Delta vert\|_2 + \lambda_{norm} \|\Delta norm\|_2 + \lambda_{edge} \mathcal{L}_{edge} + \lambda_{cons} \mathcal{L}_{cons}(norm) + \lambda_{lap} \mathcal{L}_{lap}$.

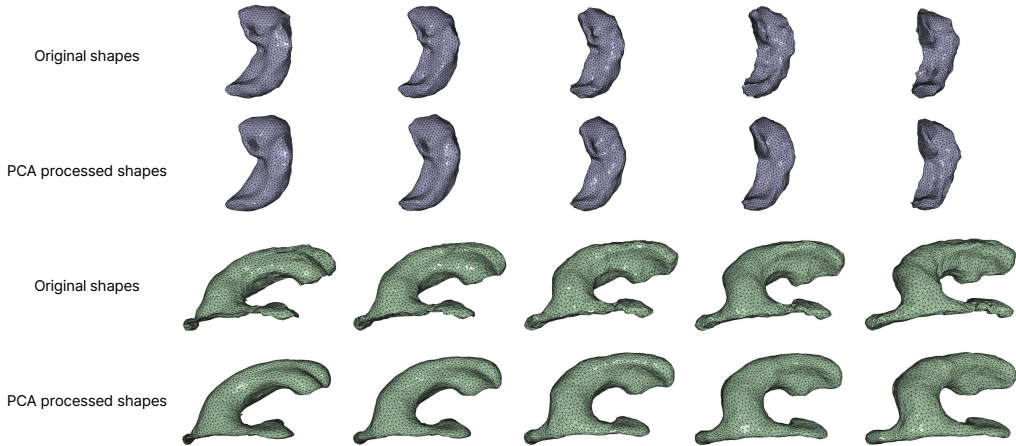
We empirically set hyperparameters for each loss as $\{\lambda_{cf}, \lambda_{pm}, \lambda_{vert}, \lambda_{norm}, \lambda_{edge}, \lambda_{cons}, \lambda_{lap}\} = \{0.5, 3, 1, 1, 1500, 1, 5\}$.

A.2 Effect of PCA on shape reconstruction



(a) Longitudinal data visualization

(b) Compactness of PCA



(c) Longitudinal shapes after PCA processing

Figure 5: Impacts of shape dimensionality reduction using PCA. (a) Visualization of vertex positions and the first PCA eigenvalue across longitudinal observations from the LBC1936 dataset. (b) Compactness measures of hippocampus and LV across varying numbers of PCA modes. (c) Qualitative comparison between original shapes and their PCA-reconstructed shapes with 150 modes.

Motivation of PCA. We adopt principal component analysis (PCA) to parameterize brain region shapes for longitudinal modeling, as described in subsection 3.1. Rather than using raw vertex-level inputs or latent codes from autoencoders, we choose PCA-based representations to facilitate reproducible and scalable research. This choice is inspired by prior work such as MANO [13], which models shape variation of hands using a low-dimensional parametric space. Building on this foundation, subsequent research has demonstrated the effectiveness of directly predicting such parameters for shape estimation tasks [5, 14, 17]. In a similar spirit, we leverage parametric modeling of brain shapes to construct compact and anatomically meaningful shape spaces to support the learning of the longitudinal shape dynamics.

Another key motivation for using PCA is its efficiency in shape representation. We visualize the vertex positions of brain shapes and the first eigenvalues across longitudinal time points in Figure 5(a). The raw vertex positions of hippocampal and LV shapes exhibit irregular deformations over time,

making it difficult to identify clear trends. This irregularity arises from the complex geometry of the original shape representation, which includes high-frequency artifacts (*e.g.*, bumpy surfaces due to image-level noise) and substantial inter-subject variability, as illustrated in Figure 5(c). In contrast, PCA-derived eigenvalues show more consistent trends across longitudinal time points, enabling stable progressive modeling of brain shape dynamics. Moreover, shapes reconstructed using PCA exhibit smoother surfaces by filtering out high-frequency noise while preserving essential geometric features (Figure 5(c)). This dimensionality reduction and smoothing effect of PCA not only simplifies the shape representation but also supports plausible modeling by providing stable inputs for capturing temporal progression in neurodegenerative disease.

Quantitative results of PCA. To quantitatively evaluate the effectiveness of PCA-based shape representation, we measured the compactness of the statistical shape models for both the hippocampus and LV, where Compactness(k) = $\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^N \lambda_i}$ for the first k modes. As depicted in Figure 5(b), the compactness reaches 0.980 for the hippocampus and 0.989 for the LV with $k = 150$ modes, indicating that over 98% of the total shape variability is preserved in both cases.

In addition, to validate the use of PCA coefficients as shape representations, we compare BrainODE performance using original vertex-based shapes V and PCA-projected shapes Λ in modeling hippocampus. As shown in Table 4, the use of PCA coefficients not only results in smoother reconstructed surfaces but also improves the prediction accuracy. The ablation study shows that using 150 components provides a favorable trade-off between reconstruction fidelity and dimensionality reduction.

Shapes	PCA Modes	Dist. (mm)
V	-	0.2642
Λ	50	0.3028
	100	0.3005
	150	0.2640
	200	0.2829
	300	0.2701

Table 4: Prediction loss in Euclidean distances (mm) when using brain shapes V as-is and using PCA coefficients Λ on hippocampal shapes of LBC1936.

A.3 Neural architecture of BrainODE

BrainODE. We design the BrainODE function f_θ as a neural network with a self-attention mechanism to model condition-injective shape dynamics. The input to f_θ consists of a shape embedding $V \in \mathbb{R}^{150}$, a scalar time $t \in \mathbb{R}^1$, and a scalar cognitive status $c \in \mathbb{R}^1$. These are concatenated into a single vector and passed through three linear projections to form the query, key, and value. The attention output is then processed by fully connected layers with GeLU activation. The output $h(\cdot)$ is a 150-dimensional vector representing the velocity in the shape space at time t .

Cognition estimator. We adopt the cognition estimator τ_θ as a simple 3D CNN architecture, followed by a prior study to diagnose Alzheimer’s disease with the hippocampus masks [9]. The input to τ_θ is a shape in voxel representation, and the output is the estimated cognitive status $c_t \in [0, 1]$.

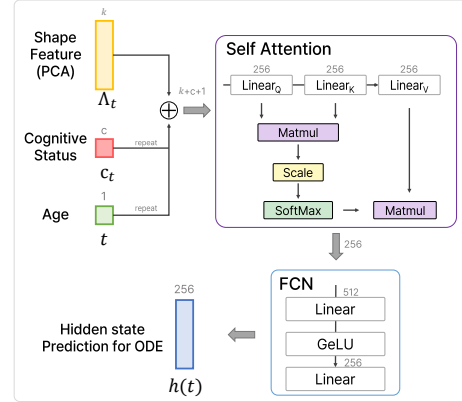


Figure 6: BrainODE architecture

A.4 Training details of BrainODE

The input to BrainODE $f_\theta(\Lambda, t, c)$ consists of PCA-reduced shape coefficients Λ , normalized age t , and cognitive status c . Each brain shape $V_i = \{v_1, v_2, \dots, v_n\}$ is represented as vertices of a triangular mesh, where the coordinates are in real-world brain sizes with units in mm. The typical spatial extents of the hippocampus and LV are approximately (50 mm, 65 mm, 35 mm) and (60 mm, 100 mm, 80 mm), respectively. These vertex coordinates are projected onto a PCA basis, and the resulting coefficients $\Lambda = \{\Lambda_1, \Lambda_2, \dots, \Lambda_n\}$ are used as shape inputs for training. To improve generalization with limited longitudinal data, we apply data augmentation via random scaling of training shapes by a factor uniformly sampled from the range [0.98, 1.02].

The temporal input t is derived by normalizing the age of subjects. Specifically, we select subjects aged between 65 and 95 years and linearly normalize their age (*i.e.*, $t = (age - 65)/30$ such that $t \in [0, 1]$). The cognitive status c is a continuous value in the range $[0, 1]$, as described in subsection 3.2.

BrainODE is trained for 100 epochs using the AdamW optimizer with a learning rate of 0.0005. For solving the neural ordinary differential equations, we employ the fourth-order Runge-Kutta (RK4) integration method. The training process minimizes the L2 loss function, as derived in Equation 3.

B Dataset details

Dataset	Subjects #	NC	AD	CONV	1st observed age	# of observ.	Intervals (yrs)
LBC1936	516	516	0	0	73.12 ± 0.79	3.59 ± 1.19	3.15 ± 0.78
AIBL	211	134	56	21	73.87 ± 5.25	3.04 ± 1.15	1.66 ± 0.57
ADNI	829	558	271	0	75.35 ± 5.81	5.05 ± 3.12	1.07 ± 0.89
OASIS2	80	65	11	4	77.42 ± 7.11	2.64 ± 0.78	2.02 ± 0.70
OASIS3	282	253	14	15	71.62 ± 4.87	2.91 ± 1.27	2.74 ± 1.50

Table 5: Composition of longitudinal datasets. NC, AD, CONV denote whose cognitive statuses are normal cognition, Alzheimer’s disease, and converted from NC to AD, respectively.

Validation data		hippo.		LV	
Time interval	Dataset	4shot	1shot	4shot	1shot
Regular	LBC	20	20	37	40
	AIBL	18	19	27	34
Irregular	AIBL + ADNI + OASIS2 + OASIS3	65	79	63	170

Table 6: Composition of validation dataset used in subsection 4.1 (Table 1 and Table 7).

B.1 Dataset Composition

To learn longitudinal brain shape dynamics, we use the LBC1936 [2] and AIBL [3] datasets as regular time interval datasets, whereas AIBL [3], ADNI [6], OASIS2 [10], and OASIS3 [8] are used as irregular time interval datasets. As shown in Table 5, each dataset varies in the number of subjects, age at the first brain image acquisition, number of observations, and intervals between scans.

The LBC1936 data are provided by the Lothian Birth Cohort 1936 Study database¹. Most participants underwent brain magnetic resonance imaging between 2008 and 2010 at an average age of 72.6 years, with follow-up scans conducted every three years over five waves. The AIBL dataset collects scans every 18 months across five waves. Since AIBL includes subjects with varied ages and provides exact MRI acquisition dates, it is utilized for both regular and irregular time interval analyses. Furthermore, because AIBL, OASIS2, and OASIS3 contain longitudinal data where cognitive status transitions from normal to Alzheimer’s disease, their shape data are used for pseudo-cognitive status shape sampling in subsection 3.2.

Validation data. The validation dataset composition is Table 6, used for the quantitative results in Table 1 and Table 7. Specifically we selected subjects whose longitudinal observations are more than four times for 4-shot prediction.

B.2 Medical priors in brain shapes.

In this section, we elaborate on the medical priors for longitudinal progression modeling of the brain, as discussed in section 1. Specifically, we analyze the volumes of the hippocampus and LV using all longitudinal datasets (ADNI, OASIS, AIBL, and LBC1936). Additionally, we visualize the rate of volume changes over time intervals to highlight the progression dynamics.

¹<https://lothian-birth-cohorts.ed.ac.uk/data-access-collaboration>

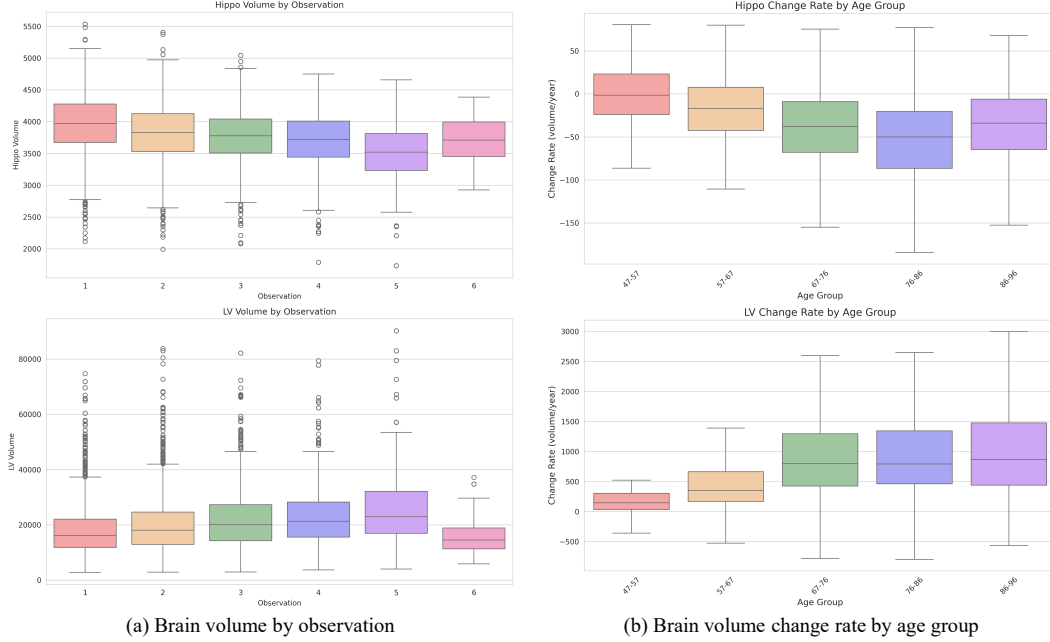


Figure 7: Illustration of brain atrophy progression. (a) Regional volumes across longitudinal observations. (b) Rate of volume changes over time intervals. The first and second rows represent the volumes of the hippocampus and LV, respectively, across all datasets (ADNI, OASIS, AIBL, and LBC1936).

Brain volume changes The human brain undergoes *brain atrophy* (i.e., the loss of brain cells) due to normal aging and neurodegenerative disorders. In particular, Alzheimer’s Disease (AD) causes more rapid brain atrophy. As brain atrophy progresses, brain volumes shrink, while the LVs enlarge as the empty space surrounded by brain subcortical structures. Figure 7(a) illustrates brain atrophy and associated volume changes across longitudinal observations for each subject in the brain MRI datasets. Hippocampal volumes show a gradual decline from observation 0 to 4, with the trend less clear at the 5th visit due to limited data availability. Conversely, LV volumes consistently increase over time, reflecting ventricular enlargement caused by the brain atrophy.

To further explore the dynamics of brain atrophy, we calculate the rate of volume changes by dividing the volume difference by the corresponding time intervals (e.g., 1.5 years for AIBL, 3 years for LBC1936, and respective years for irregular time intervals). This rate, expressed in $mm^3/year$, quantifies the speed of atrophy or enlargement. Figure 7(b) visualizes these rates across age groups, defined based on the dataset distributions. The hippocampus shows a higher rate of volume reduction in older age groups compared to the 47–57 group, whereas LV enlargement accelerates with age. Notably, since older brains are smaller in volume, a similar absolute rate of volume loss implies relatively more severe atrophy in later life stages (e.g., 76–86 vs. 86–96 age groups). These trends align with the medical priors, confirming that older age exacerbates the progression of atrophy.

To assess the impact of cognitive status, Figure 8 compares the rate of volume changes between normal cognition (NC) and AD in the same age groups. We visualize the change rate of the left and right brain regions separately. AD (red) exhibits a significantly faster and more irregular rate of volume change than NC (blue). For example, in the 67–76 age group, the LV volume reduction rate is approximately $-700 mm^3/year$ for NC, compared to $-1600 mm^3/year$ with greater variability in AD. This heterogeneous pathological progression of AD presents challenges for modeling as it requires capturing complex and non-linear patterns that diverge from the more stable trends observed in NC.

C Additional analysis and visualization

In this section, we comprehensively evaluate longitudinal brain shape prediction on multiple perspectives. We first present additional experimental results from subsection 4.1, including the qualitative

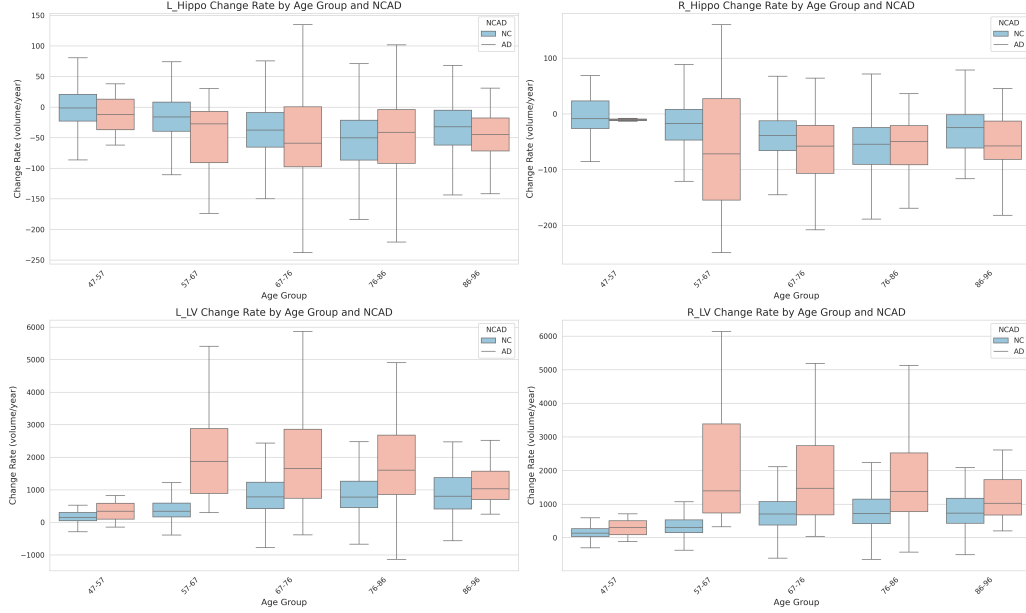


Figure 8: Rate of volume changes over time intervals for normal cognition (NC) and Alzheimer’s Disease (AD). The age groups are split based on the data distribution from ADNI, OASIS, AIBL, and LBC1936. NC and AD groups are visualized in blue and red, respectively

566 results of the 1-shot prediction in subsection C.1, the effectiveness of the pseudo-cognitive status
567 embedding in subsection C.2, and BrainODE’s condition fidelity in modeling Alzheimer’s Disease
568 (AD) in subsection C.3. Furthermore, we examine the subject inconsistency inherent in image gener-
569 ative approaches for brain region progression modeling compared to deformation-based methods in
570 subsection C.4. Finally, we evaluate the long-range shape prediction capabilities of baseline methods
571 and BrainODE in subsection C.5.

572 C.1 Additional analysis of experiments

573 Qualitative evaluation of 1-shot prediction.

574 We present qualitative results for brain shape modeling using 1-shot prediction in Figure 9. Unlike the
575 4-shot setting in Figure 2, which leverages multiple time points to guide the deformation trajectory,
576 the 1-shot setting lacks temporal deformation information. As a result, conventional extrapolation
577 methods are excluded.

578 To implement prediction on arbitrary time points using ShapeFlow [7], which does not support
579 irregular time intervals, we train the model on the AIBL [3] dataset using fixed 1.5-year intervals.
580 The starting time point is set to $t = 0$, and 1.5 years is mapped to $t = 1$. Consequently, future time
581 targets are normalized relative to this interval (*e.g.*, a 2-year prediction corresponds to $t = 2/1.5$).

582 Other baseline methods that can handle irregular time intervals are also included for comparison.
583 This setup allows us to evaluate the ability of each method to predict future shapes under varying
584 temporal gaps. Notably, addressing irregular time intervals is critical in longitudinal medical imaging,
585 as patient visit schedules often vary. Robust modeling across inconsistent time gaps is essential for
586 accurately capturing disease progression.

587 Figure 9 visualizes qualitative results of 1-shot prediction for hippocampus and LV across varying
588 ages and cognitive statuses, including normal cognition (NC) and Alzheimer’s disease (AD). Overall,
589 BrainODE outperforms baseline methods in modeling both brain regions. Specifically, for the
590 hippocampus, both BrainODE and ShapeFlow achieve small Euclidean distances from the ground
591 truth. However, BrainODE uniquely excels in LV prediction, demonstrating the lowest errors.

592 Regarding the baseline results, ShapeFlow accurately predicts hippocampus shapes but struggles to
593 model the LVs, showing significant errors (> 4 mm) in several regions. LatentODE [15] produces

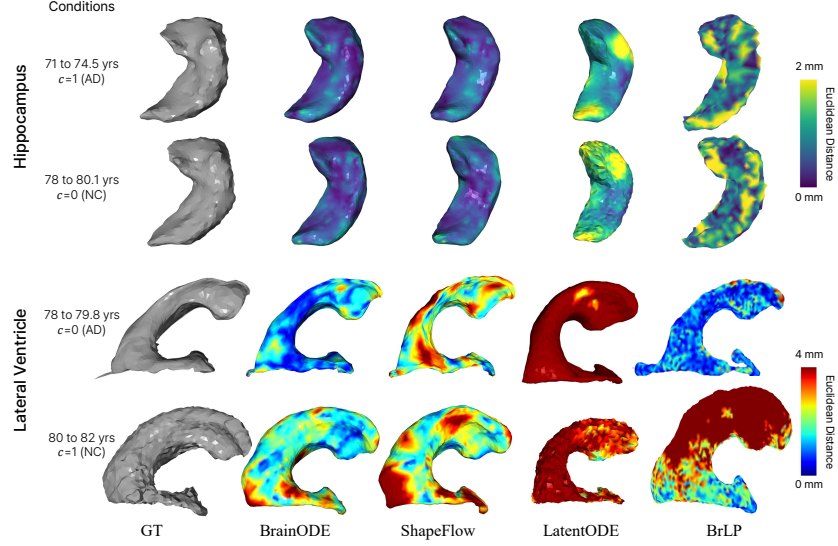


Figure 9: Qualitative results of 1-shot prediction for hippocampus and LV shapes. The color map indicates Euclidean distance errors (up to 2 mm for hippocampus, 4 mm for LV) between ground truth (GT) and predicted shapes.

incorrect LV shapes by altering their topology and fails to capture subject-specific features, generating nearly identical shapes regardless of input conditions. This suggests that LatentODE struggles to model the prior distribution of brain progression dynamics using autoencoders, leading to statistically plausible yet trivial solutions (*e.g.*, converging to average shapes) that disregard the given inputs. Similarly, BrLP distorts the anatomical features of the input shapes. For instance, over-enlargements are observed on the occipital horn, frontal horn, and body of the LV in NC subjects, and topological changes appear in occipital horn of the AD subject. In contrast, despite the challenges of modeling brain progression with irregular time intervals, BrainODE consistently demonstrates superior performance, validating its robustness and accuracy.

Shape prediction accuracy by cognitive status.

Further details of subsection 4.1 are presented in Table 7. This section reports shape prediction performance on irregular longitudinal datasets from ADNI, OASIS, and AIBL, grouped by cognitive statuses of normal cognition (NC) and Alzheimer’s disease (AD). Our BrainODE achieves similar accuracy in both groups with average deviations around 0.1 mm. Notably, despite the heterogeneous pathology and high deviations in the hippocampus and LV of the AD subjects revealed in subsection B.2, our BrainODE demonstrates well-generalized predictions for both NC and AD subjects.

cognition	Hippocampus		LV	
	4-shot	1-shot	4-shot	1-shot
NC ($c = 0$)	0.540	0.466	2.489	1.728
AD ($c = 1$)	0.588	0.585	2.277	1.691
Total	0.543	0.492	2.469	1.711

Table 7: Quantitative evaluation of shape prediction performance by cognitive status c in Euclidean distance (mm) on ADNI, OASIS, AIBL test sets with irregular time points.

C.2 Sampling longitudinal shapes with intermediate timepoints

In this section, we explore the effects of incorporating pseudo-cognitive status embeddings and generating shapes on intermediate time points to enable progressive longitudinal modeling. Specifically, we investigate how this approach impacts the accuracy and stability of shape predictions. To this end, we compare the qualitative results of shapes generated every 3 years over a 12-year period with those produced in a single step (*i.e.*, 1-hop prediction).

Figure 10 illustrates the results, showing the deformation of hippocampus and LV shapes from age 73 to 85, along with the Euclidean distance maps from the initial shape at age 73. Predicting brain shapes over long intervals in a single step is challenging for BrainODE, as such large temporal gaps are not seen during training, where the average interval between the first and last acquisitions is

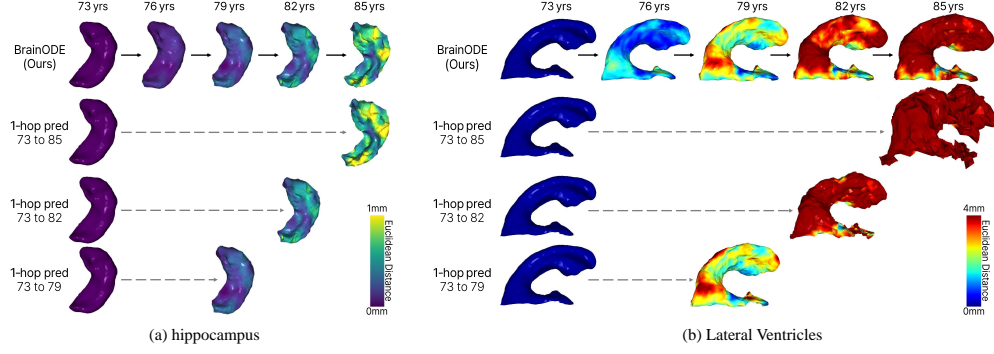


Figure 10: Qualitative results of sampling longitudinal shapes with intermediate time points for (a) hippocampus and (b) LVs.

626 4.08 ± 2.84 years. To address this, BrainODE adopts an iterative prediction strategy, forecasting
 627 future shapes through intermediate time steps, as shown in the first row of the figure.

628 Direct prediction over a 12-year interval results in unrealistic deformations and surface artifacts;
 629 especially, the LV surface is almost entirely destroyed. In contrast, when intermediate predictions
 630 are made every 3 years, the final 12-year prediction maintains anatomical consistency and surface
 631 smoothness. This experiment highlights the effectiveness of progressive pseudo-shape sampling in
 632 modeling longitudinal brain progression.

633 C.3 Tendency of shape changes by age and cognitive state

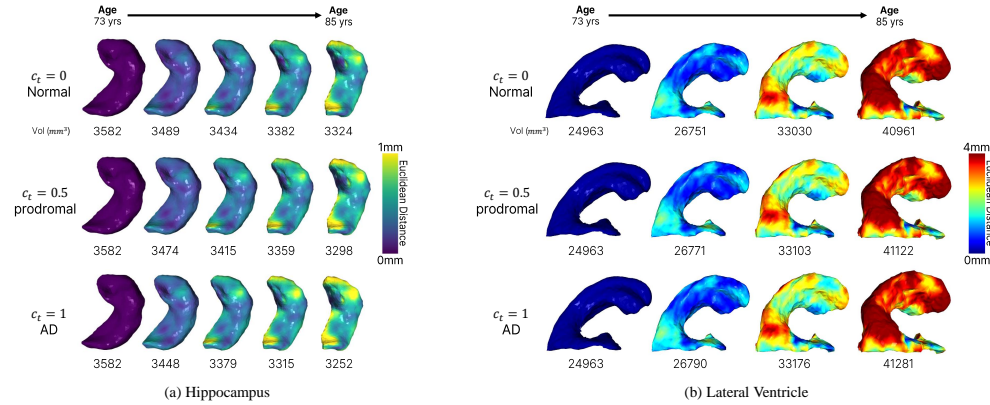


Figure 11: Longitudinal shape predictions with varying cognitive status $c = \{0, 0.5, 1\}$ every 3 years for hippocampus and LV. The number below each shape denotes the volume of the predicted shape.

634 To further demonstrate the *condition-injectivity* of BrainODE,
 635 we visualize shape predictions under varying cognitive statuses
 636 $c = \{0, 0.5, 1.0\}$ in Figure 11. BrainODE predicts larger deformations
 637 in both the hippocampus and LVs as cognitive status declines,
 638 consistent with clinical progression patterns observed in longi-
 639 tudinal data (subsection B.2). For instance, the shape predicted
 640 with intermediate status $c = 0.5$ shows deformations that lie be-
 641 tween those of normal cognition and Alzheimer’s disease, as also
 642 reflected in their estimated volumes.

643 Beyond qualitative visualization, we analyze volumetric trends
 644 across varying cognitive statuses for 100 baseline hippocampal
 645 shapes in Figure 12 at the age of 65. Even though explicit medical
 646 priors is not injected during training, BrainODE captures clinically

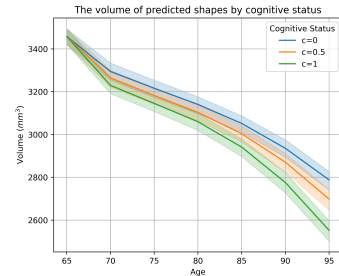


Figure 12: Brain Atrophy tendency by cognitive status c modeled by BrainODE.

plausible volume trajectories by leveraging the spatiotemporal supervision from longitudinal datasets. This supports the ability of BrainODE to encode implicit medical priors through learning dynamics across both age t and cognitive status c .

C.4 Subject inconsistency of generative approaches

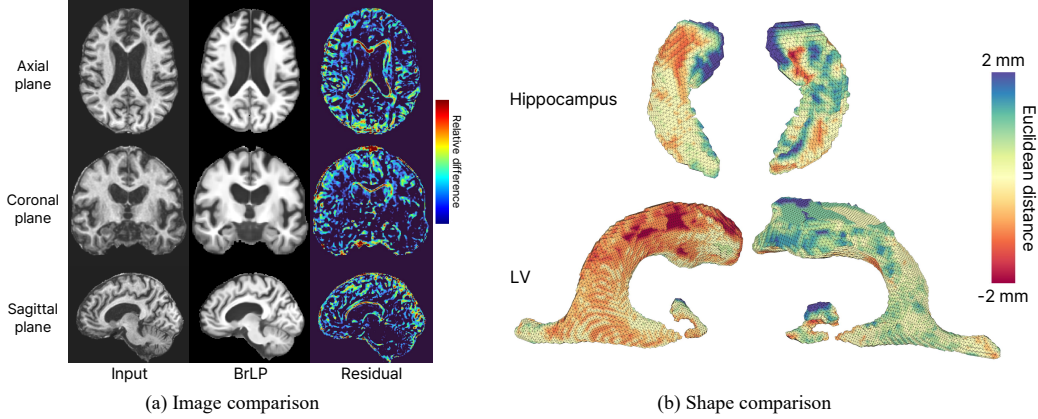


Figure 13: Subject inconsistency of the BrLP [12] generative model. (a) Image comparison shows the input brain MRI scan (axial, coronal, sagittal planes) and the residual difference after BrLP generation with $\Delta t = 0$. (b) Shape comparison highlights Euclidean distance errors (up to ± 2 mm) between the input and generated hippocampus and LV shapes.

In this study, we examine two primary approaches for modeling longitudinal progression in neurodegenerative diseases: generative models and deformation-based methods, which exhibit distinct characteristics. Generative approaches, such as BrLP [12], model medical priors using the distribution of given data and its associated conditions (*e.g.*, age and cognitive status). During inference, these models generate statistically plausible samples from the learned distribution. However, as discussed in subsection 4.1, we hypothesize that this approach is unsuitable for longitudinal progression modeling of medical data, as it is challenging to ensure spatial consistency of each subject (*i.e.*, maintaining the sampled data’s identity). To test this, we conducted a preservation experiment using the BrLP generative model by inputting a brain MRI scan with no time difference ($\Delta t = 0$) and measuring the differences between the input and generated outputs in both image and shape spaces, as depicted in Figure 13.

In the image-level comparison (Figure 13(a)), although the output is expected to be identical to the input, noticeable differences are observed in high-frequency details and anatomical structures such as the LV, as highlighted by their residuals in the color map. We further visualize shapes of the hippocampus and LVs, which are of two key brain regions to determine the cognitive status, in the shape-level comparison (Figure 13(b)). The Euclidean distance is calculated with the input subject’s shapes, indicating significant deviations in the hippocampus and LV shapes. In addition, the LV has a seamless structure in the brain, however, BrLP generates a disconnected LV due to limited voxel grid resolution. These inconsistencies could compromise longitudinal diagnosis, potentially perturbing the AD decision boundary of deterministic classification models sensitive to subtle changes, as reported in [9]. This highlights a fundamental limitation of generative models in maintaining subject-specific fidelity. In contrast, deformation-based approaches such as BrainODE offer improved consistency by preserving anatomical identity when $\Delta t = 0$, making them more suitable for reliable longitudinal modeling.

C.5 Longitudinal shape prediction capacity

From a practical perspective, an important question arises in longitudinal modeling: how far into the future can models reliably predict brain shape progression while maintaining anatomical consistency? To investigate this, we evaluated the longitudinal prediction capabilities of five models—ShapeFlow, Extrapolation, BrLP, LatentODE, and BrainODE—using the LBC1936 dataset, as shown in Figure 14.

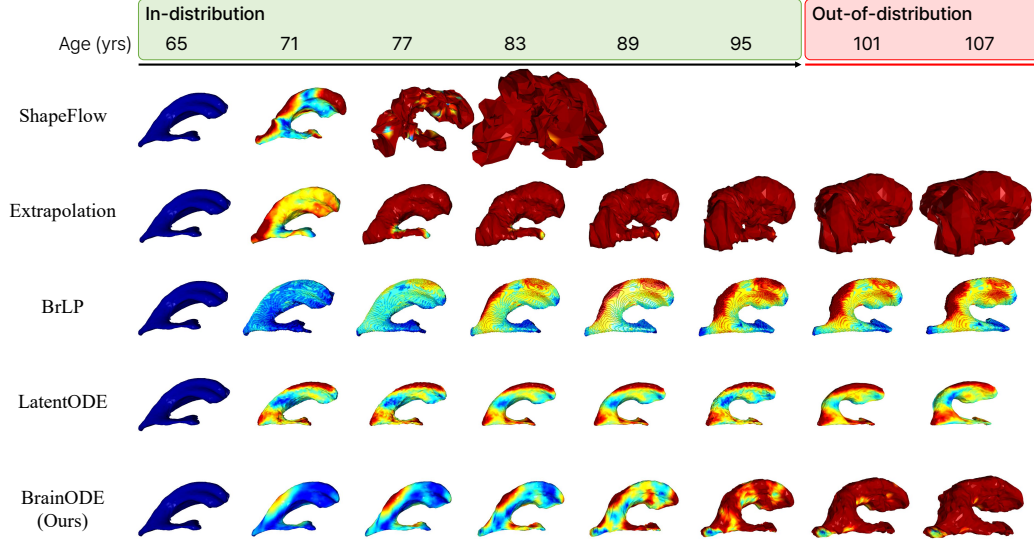


Figure 14: Comparison of longitudinal shape prediction over time for the hippocampus and LV. Each column represents a different model, showing shapes predicted from ages 65 to 107 at uniform intervals. Colors indicate the Euclidean distance from the first shape at age 65.

The figure presents the predicted hippocampus shapes of a single subject spanning ages 65 to 95 (within the training distribution) and extending to 107 years (beyond the training distribution), demonstrating each model’s ability to predict long-term brain shape progression.

Specifically, ShapeFlow fails to predict beyond 89 years, as the brain shapes "explode", making visualization infeasible. Extrapolation, while able to generate shape predictions, shows progressive collapse when forecasting further beyond the two input time points (*e.g.*, 65 and 71 years), with noticeable distortions at 95 years. BrLP appears plausible at a glance, but as discussed in subsection C.4, exhibits subject inconsistency; the predicted shape at 71 years already shows changes in overall geometry and new appearances, not in the shape at 65 years. LatentODE, consistent with the main experimental results, struggles with prior modeling within its VAE framework, repeatedly predicting similar shapes (*e.g.*, minimal variation in LV shape across all ages) due to training difficulties. In contrast, BrainODE demonstrates robust performance within the in-distribution age, progressively transforming shapes while preserving anatomical geometry features, such as the curvature of the hippocampus and the expansion patterns of the LV, even up to 95 years. This highlights superior capacity of BrainODE for long-term longitudinal prediction compared to other models.

D Limitations and Future Work

In this work, we utilize two cognitive status groups: normal cognition (NC) and Alzheimer’s disease (AD). Although some datasets include mild cognitive impairment (MCI) as a separate category and intermediate cognition between NC and AD, we excluded MCI from BrainODE modeling due to its ambiguous clinical definition and variability in progression (*e.g.*, assigning $c := 0.5$ for MCI seems straightforward yet cannot reflect severity in the values). In future work, we aim to extend BrainODE to incorporate NC, MCI, and AD by establishing more robust and discriminative criteria for cognitive status.

For subjects with multiple longitudinal observations ($n \geq 1$), BrainODE can generate n predicted shapes corresponding to each time point. To evaluate n -shot prediction performance, we simply average these predicted shapes. Developing more advanced aggregation strategies—such as attention-based or time-aware mechanisms—remains an important direction for future research to better capture longitudinal progression.

As a pioneering model for brain shape prediction under varying cognitive conditions, BrainODE introduces a pseudo-cognitive status embedding that enables intermediate shape prediction and the

corresponding cognition estimation. In this work, we primarily estimate cognitive status using hippocampal shapes, guided by prior research [9, 16]. In future work, we plan to explore BrainODE and cognition estimator to incorporate additional brain regions.

References

- [1] Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, Juan Eugenio Iglesias, et al. Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical image analysis*, 86:102789, 2023.
- [2] Ian J Deary, Alan J Gow, Michelle D Taylor, Janie Corley, Caroline Brett, Valerie Wilson, Harry Campbell, Lawrence J Whalley, Peter M Visscher, David J Porteous, et al. The lothian birth cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC geriatrics*, 7:1–12, 2007.
- [3] Kathryn A Ellis, Cassandra Szoeka, Ashley I Bush, David Darby, Petra L Graham, Nicola T Lautenschlager, S Lance Macaulay, Ralph N Martins, Paul Maruff, Colin L Masters, et al. Rates of diagnostic transition and cognitive change at 18-month follow-up among 1,112 participants in the australian imaging, biomarkers and lifestyle flagship study of ageing (aibl). *International Psychogeriatrics*, 26(4):543–554, 2014.
- [4] Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. Point2mesh: A self-prior for deformable meshes. *ACM Trans. Graph.*, 39(4), 2020.
- [5] Lin Huang, Jianchao Tan, Jingjing Meng, Ji Liu, and Junsong Yuan. Hot-net: Non-autoregressive transformer for 3d hand-object pose estimation. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3136–3145, 2020.
- [6] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
- [7] Chiyu Jiang, Jingwei Huang, Andrea Tagliasacchi, and Leonidas J Guibas. Shapeflow: Learnable deformation flows among 3d shapes. *Advances in Neural Information Processing Systems*, 33:9745–9757, 2020.
- [8] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medrxiv*, pages 2019–12, 2019.
- [9] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, Mingqing Xu, Alzheimer’s Disease Neuroimaging Initiative, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer’s disease. *Neuroimage*, 208:116459, 2020.
- [10] Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. *Journal of cognitive neuroscience*, 22(12):2677–2684, 2010.
- [11] Wonjung Park, Maria del C Valdés Hernández, Jaeil Kim, Susana Muñoz Maniega, Fraser N Sneden, Karen J Ferguson, Mark E Bastin, Joanna M Wardlaw, Simon R Cox, and Jinah Park. Ai-based deformable hippocampal mesh reflects hippocampal morphological characteristics in relation to cognition in healthy older adults. *NeuroImage*, 310:121145, 2025.
- [12] Lemuel Puglisi, Daniel C Alexander, and Daniele Ravì. Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 173–183. Springer, 2024.

- 757 [13] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and
758 capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*,
759 36(6), November 2017.
- 760 [14] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. Frankmocap: A monocular 3d whole-body
761 pose estimation system via regression and integration. In *IEEE International Conference on*
762 *Computer Vision Workshops*, 2021.
- 763 [15] Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. Latent ordinary differential equations
764 for irregularly-sampled time series. *Advances in neural information processing systems*, 32,
765 2019.
- 766 [16] Ingrid S van Maurik, Stephanie J Vos, Isabelle Bos, Femke H Bouwman, Charlotte E Teunissen,
767 Philip Scheltens, Frederik Barkhof, Lutz Frolich, Johannes Kornhuber, Jens Wiltfang, et al.
768 Biomarker-based prognosis for people with mild cognitive impairment (abide): a modelling
769 study. *The Lancet Neurology*, 18(11):1034–1044, 2019.
- 770 [17] Taeyun Woo, Wonjung Park, Woohyun Jeong, and Jinah Park. A survey of deep learning
771 methods and datasets for hand pose estimation from hand-object interaction images. *Computers*
772 *& Graphics*, 116:474–490, 2023.