

Progression as Latent Drift: Generative Forecasting of Slow-Evolving Pathologies

Yuxiang Feng^{1,3*†}, Juncheng Wang^{2*}, Chao Xu^{3,4}, Wenlong Hou²
Huihan Wang², Yijie Qian^{1,3}, Yang Liu^{3,4}, Baigui Sun^{3,4}
Yong Liu^{1†}, and Shujun Wang^{2†}

¹ Zhejiang University, Hangzhou, China

² The Hong Kong Polytechnic University, Hong Kong, China

³ IROOTECH TECHNOLOGY, China

⁴ Wolf 1069 b Lab, Sany Group, China

* Equal contribution. † Corresponding authors.

{fengyx@zju.edu.cn, wjc2830@gmail.com}^{*}, {yongliu@iipc.zju.edu.cn,
shu-jun.wang@polyu.edu.hk}[†]

Abstract. Forecasting the future anatomy of slow-evolving neurodegenerative diseases could enable earlier, more targeted intervention and improve clinical trial design, but it remains challenging because true progression signals are subtle in longitudinal MRI. In this low-signal regime, transferring modern generative sequence models directly is unreliable: training is dominated by stable baseline anatomy and confounded by dense, sample-specific nuisance variation. We first provide a theoretical analysis that explains these failures through two modes. Identity collapse occurs when optimization is driven toward reproducing the current anatomy, which prevents the model from learning faint temporal change. The continuous interpolation trap arises when standard smooth networks cannot separate localized biological drift from pervasive noise, which leads to spurious changes that diffuse across the volume. To address both issues, we propose Latent Drift, a progressive generative framework that learns change in a compressed semantic representation rather than synthesizing full-resolution anatomy. This design removes pixel-level identity from the prediction target and concentrates model capacity on progression-relevant dynamics. We further apply Finite Scalar Quantization to the learned change representation, which suppresses small, high-frequency nuisance fluctuations while preserving consistent structural drift. Experiments on longitudinal 3D brain MRI show that Latent Drift improves patient-specific neuro-forecasting over diffusion and autoregressive transformer baselines across generative fidelity and clinically relevant evaluation metrics.

Keywords: Slow-Evolving Pathologies · Generative Model · Brain Simulation

1 Introduction

Dynamic simulation of organ evolution [6, 10, 46] is essential for forecasting disease trajectories, particularly for neurodegeneration and other slow-evolving

[‡] This work was conducted in collaboration with IROOTECH TECHNOLOGY.

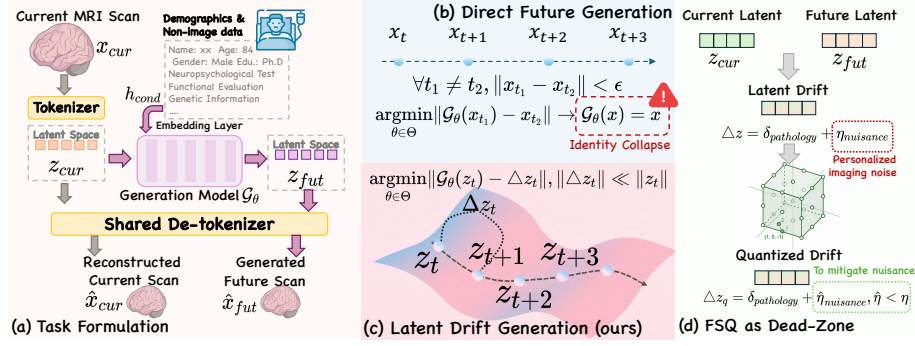


Fig. 1: Progression as Latent Drift. (a) **Task Formulation**: Autoregressive forecasting of future MRI states in a compressed semantic latent space. (b) **Direct Future Generation**: Predicting the absolute future state induces *Identity Collapse*, as stationary background anatomy overwhelms microscopic biological changes. (c) **Latent Drift Generation (Ours)**: Predicting the temporal residual (Δz) bypasses this collapse, isolating the true pathological trajectory. (d) **FSQ as a Topological Dead-Zone**: To break the *Continuous Interpolation Trap* caused by dense imaging noise, Finite Scalar Quantization (FSQ) serves as a non-Lipschitz filter that annihilates nuisance variations while preserving sparse semantic drift.

brain pathologies that progress along largely irreversible courses. Because neuronal loss and downstream structural changes (e.g., cortical thinning and ventricular enlargement) cannot be recovered [1, 4], substantial tissue damage may already be present by the time symptoms become clinically apparent [3, 11]. This delayed observability narrows the effective window for intervention and complicates the development and evaluation of disease-modifying therapies, which increasingly depend on identifying who will deteriorate, where, and how fast [2, 32, 56].

Generative forecasting predicts a patient’s future anatomical state and simulates the personalized morphological trajectory of disease. Localized anticipation of progression can support earlier, more targeted care for conditions such as Alzheimer’s and Parkinson’s disease, where structural decline unfolds over years [36, 42]. It can also enrich clinical trials by prioritizing participants likely to show measurable progression, and reduce caregiving burden through more proactive management before irreversible damage accumulates [48].

Recent generative sequence models synthesize audio [18, 50–52] and visual [5, 8, 25, 29, 30, 49] content well, which makes them a natural starting point for forecasting future brain anatomy [7, 24, 35, 43], as shown by Fig. 1-(a).

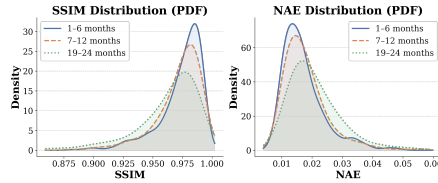


Fig. 2: Statistic similarity between current and future states.

However, as shown in Fig. 2, forecasting neurodegeneration is different because current and future MRI states are statistically highly similar, making the temporal progression signal far weaker. Unlike natural videos with salient motion, neurodegeneration unfolds over long time scales [19, 28]: across a typical one-year interval, disease-related morphological change may be under 1% of the total volumetric variance in a 3D MRI scan [33]. In this low-signal regime, naive transfer is unreliable, as training is dominated by static anatomy or by nuisance variation resembling scanner noise rather than the subtle biological progression of interest [14, 53].

To clarify this mismatch, our first contribution is a mathematical analysis showing that directly applying state-of-the-art generative sequence models to neurodegenerative forecasting is structurally prone to failure. The analysis yields two characteristic failure modes that expose intrinsic limitations of existing approaches. First, **identity collapse** occurs when optimization is dominated by the much larger baseline anatomy, pushing the model toward an approximate identity mapping that reproduces static neuroanatomy instead of learning faint temporal drift, as shown by Fig. 1-(b). Second, **the continuous interpolation trap** arises under dense, sample-specific nuisance variation. For Lipschitz-continuous networks, sparse semantic progression cannot be separated cleanly from pervasive noise. The learned predictor then behaves as a spurious interpolator that diffuses variance across the full volume instead of recovering localized neurodegenerative change [15, 45]. Therefore, *How can we constrain a generative model to reliably predict subtle, localized atrophy trajectories, instead of memorizing baseline anatomy or amplifying scanner-dependent noise?*

In this paper, we address these two failure modes with a progressive generative framework designed to separate the disease-relevant progression signal from both dominant background anatomy and nuisance variability. First, to mitigate the dominance of baseline anatomy and prevent identity collapse, we project each MRI scan into a compressed semantic latent space and train an autoregressive model to predict **the latent drift** between the current and future representations, rather than **reconstructing the entire future image**. By removing pixel-level identity as the modeling target, the model is encouraged to allocate its capacity to learning the temporal change that reflects disease progression, illustrated by Fig. 1-(c). Second, to reduce the influence of nuisance variation and avoid spurious interpolation, we apply Finite Scalar Quantization (Fig. 1-(d)) to the latent change representation. This step acts as a **topological dead-zone filter**: small, high-frequency variations are mapped to zero, while consistent structural changes associated with disease progression stay above the quantization threshold and are preserved for forecasting.

We evaluate our Latent Drift framework on longitudinal 3D brain MRIs. By bypassing Identity Collapse and filtering nuisance noise, our approach improves neuro-forecasting over prior diffusion and autoregressive baselines. Our main contributions are as follows.

- **Task formalization and failure analysis:** We cast forecasting of slow-evolving pathologies as conditional generative synthesis and identify two fundamental optimization failures: **identity collapse** from dominant stationary anatomy and the **continuous interpolation trap** from dense nuisance variability.
- **Latent Drift framework:** We introduce a progressive generative architecture that predicts progression in a compressed semantic space, disentangling biological change from baseline anatomy and amplifying the sparse pathological signal.
- **FSQ mechanism:** We analyze Finite Scalar Quantization as a **topological dead-zone filter** that suppresses high-frequency nuisance noise while preserving structurally consistent drift, mitigating the limitations of smooth predictors.
- **Empirical performance:** We achieve state-of-the-art patient-specific 3D MRI forecasting, outperforming diffusion and autoregressive transformer baselines on both generative fidelity and clinically relevant metrics.

2 Preliminaries

2.1 Problem Formulation

Let $\mathcal{D} = \{(x_{cur}^{(i)}, x_{fut}^{(i)}, c^{(i)}, \Delta t^{(i)})\}_{i=1}^N$ be a dataset of longitudinal 3D brain MRIs. Given a baseline scan $x_{cur} \in \mathbb{R}^{H \times W \times D}$, patient clinical metadata c , and a time horizon Δt , the objective is to learn a conditional generative model $p_\theta(x_{fut} \mid x_{cur}, c, \Delta t)$ to forecast the future anatomical state x_{fut} .

2.2 Baseline: Latent Generative Forecasting

Due to the computational intractability of modeling high-resolution 3D volumes directly in pixel space, the standard baseline paradigm employs a three-stage latent generative framework.

- **Spatial Compression:** A pre-trained encoder $\mathcal{E}$ maps the input volumes into a compact, lower-dimensional latent space: $z_{cur} = \mathcal{E}(x_{cur})$ and $z_{fut} = \mathcal{E}(x_{fut})$.
- **Latent Generation:** The clinical variables and time horizon are projected into dense conditioning embeddings, $h_{cond} = \text{MLP}(c, \Delta t)$. A sequence model $\mathcal{G}_\theta$ (e.g., an autoregressive transformer) is then trained to directly synthesize the future latent state:

$$\hat{z}_{fut} = \mathcal{G}_\theta(z_{cur}, h_{cond}). \quad (1)$$

- **Reconstruction:** The forecasted MRI is decoded back to the pixel domain via a pre-trained spatial decoder $\mathcal{D}$, yielding $\hat{x}_{fut} = \mathcal{D}(\hat{z}_{fut})$.

3 The Pathologies of Decomposed Generative Forecasting

Having established the standard latent generative baseline, we analyze the optimization dynamics of forecasting slow-evolving temporal systems. To understand why standard architectures fail on Alzheimer’s progression, we decompose the future latent state into three signals:

$$z_{fut} = z_{cur} + \delta_{pathology} + \eta_{nuisance}, \quad (2)$$

where z_{cur} is the dominant stationary anatomy, $\delta_{pathology}$ is the sparse biological progression, and $\eta_{nuisance}$ is the dense, sample-personalized imaging noise. In this section, we show how the properties of these terms induce optimization pathologies that make standard continuous generation insufficient.

3.1 Source Dominance and the Identity Trap

Our first hypothesis addresses the relative magnitudes of the decomposed terms: because the structural change over a typical clinical horizon is microscopic,

the magnitude of the stationary background vastly exceeds the biological signal ($\|z_{cur}\| \gg \|\delta_{pathology}\|$). We hypothesize that when a generative sequence model $\mathcal{G}_\theta$ is optimized to directly synthesize the absolute future state (z_{fut}), this magnitude imbalance causes the stationary background to overwhelmingly dominate the training gradients, trapping the model in an identity function ($f(x) \approx x$).

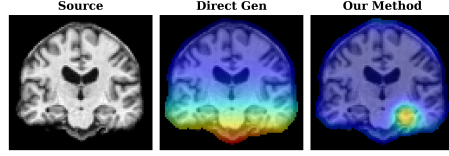


Fig. 3: Visualization of gradient-based signal differences between current and predicted future states.

Empirical Validation: To empirically validate this source dominance, we trained a baseline continuous sequence model to directly predict z_{fut} conditioned on patient metadata and time horizon. To trace the optimization focus, we visualized the spatial gradients of the objective function with respect to the generated future state during training.

As shown in Fig. 3, the gradient analysis shows a clear difference in how capacity is allocated. When generating z_{fut} directly, the training gradients spread across the spatial volume, indicating that the network spent its capacity minimizing reconstruction error on static anatomy such as the skull and unaffected hemispheres, and largely ignored the localized progression. When the target was shifted to the temporal difference (Δz), the gradients localized around regions of true change such as the hippocampus. This confirms that predicting the absolute state induces Identity Collapse, and that a successful architecture must remove z_{cur} from the objective to isolate the progression content. Analogous spatial-temporal decoupling has proven effective for motion modeling in natural video [31], though our setting demands isolating a far sparser biological signal.

3.2 The Continuous Interpolation Trap

Shifting the predictive target to the temporal residual ($\Delta z = \delta_{pathology} + \eta_{nuisance}$) bypasses Identity Collapse, but it exposes the model to a second pathology. The dense, sample-personalized nature of $\eta_{nuisance}$ perturbs the convergence of continuous neural networks and prevents recovery of the true biological trajectory.

We prove this using statistical learning theory. The biological drift δ is anatomically sparse, supported on a localized, low-dimensional manifold. The imaging noise η is a high-frequency perturbation distributed across the entire latent dimension. When a continuous sequence model fits this composite signal, it behaves as a spurious interpolator.

We formalize this failure mode in the following theorem, demonstrating that continuous, Lipschitz-bounded models are mathematically incapable of isolating sparse semantic drift when corrupted by dense, sample-personalized noise.

Theorem 1 (Continuous Interpolation Trap(Informal, see Appendix for formal version)). *Let $\Delta z(x) = \delta(x) + \eta(x)$, where $\delta(x) \in \mathbb{R}^d$ is sparse ($\|\delta(x)\|_0 \leq s \ll d$) and $\eta(x)$ is dense, zero-mean nuisance noise with $\|\eta(x)\|_\infty \leq \varepsilon$ and non-zero entries in all coordinates.*

Let f_θ be an L -Lipschitz model trained by ERM to fit Δz .

Then, even if f_θ interpolates the training data,

$$\mathbb{E}\|f_\theta(x) - \delta(x)\|_2^2 \gtrsim \mathbb{E}\|\eta(x)\|_2^2, \quad (3)$$

and for unseen x , $f_\theta(x)$ is generically dense.

Hence, no Lipschitz-continuous predictor trained on Δz can recover the sparse support of δ .

Implications for Forecasting: Theorem 1 establishes a mathematical limit on continuous generative forecasting. Because modern neural networks are Lipschitz-continuous, minimizing expected error over the dense $\eta_{nuisance}$ distribution forces the model to spread small, continuous variations across the entire predicted latent volume. The resulting generation smears spurious variance across the volume, which compromises the structural fidelity of the predicted Alzheimer’s trajectory.

In summary, the empirical trap of z_{cur} and the theoretical trap of $\eta_{nuisance}$ together impose an architectural requirement: *the generative framework must target the temporal drift to escape Identity Collapse, and it must have a non-Lipschitz, non-linear thresholding property to break the continuous interpolation trap.*

4 Our Approach: Progression as Latent Drift

Our pathology analysis imposes two constraints at once: the model must escape source dominance while also breaking the continuous interpolation trap. To meet both, we introduce a two-stage generative framework, Progression Latent Drift,

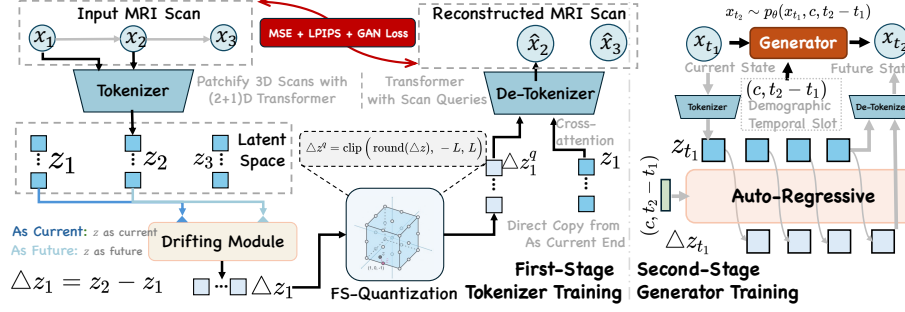


Fig. 4: The Progression Latent Drift Framework. Our two-stage architecture explicitly disentangles sparse biological progression from stationary anatomy and dense imaging noise. **First-Stage (Latent Drift Tokenizer):** Longitudinal MRI scans are patchified and compressed into continuous latent representations (z). To escape Identity Collapse, a Drifting Module computes the temporal residual ($\Delta z_1 = z_2 - z_1$). Finite Scalar Quantization (FSQ) discretizes this continuous drift, acting as a topological dead-zone filter to strictly suppress non-semantic nuisance noise. The future scan is then reconstructed by fusing the quantized drift (Δz_1^q) with the baseline latent (z_1) via cross-attention. **Second-Stage (Generator Training):** With the biological trajectory successfully isolated, an autoregressive sequence model learns to forecast the discrete drift tokens, conditionally guided by the baseline anatomy (z_{t_1}), patient demographics, and the target temporal horizon ($t_2 - t_1$).

that separates the semantic biological trajectory from the stationary background and the dense imaging noise. The overall pipeline is shown in Fig. 4.

Architecture overview. The framework combines a residual tokenization stage with an autoregressive sequence model. Rather than mapping $x_{cur} \rightarrow x_{fut}$ directly, we train a Latent Drift Tokenizer to isolate and discretize the temporal difference between scans. We then train a Decoder-Transformer to forecast this discrete progression signal, conditioned on the patient’s baseline state and clinical metadata.

4.1 The Latent Drift Tokenizer (Escaping Identity Collapse)

To address the Identity Collapse caused by the large spatial overlap of x_{cur} , we use an architectural bottleneck that prevents the network from reconstructing the stationary anatomy from the temporal target.

Given a longitudinal pair (x_{cur}, x_{fut}) , a tokenizer $\mathcal{E}$ first compresses the volumes into dense continuous representations: $z_{cur} = \mathcal{E}(x_{cur})$ and $z_{fut} = \mathcal{E}(x_{fut})$. Instead of quantizing these absolute states independently, we compute the continuous latent drift:

$$\Delta z_{raw} = z_{fut} - z_{cur}. \quad (4)$$

To sanitize this signal of dense, sample-personalized noise ($\eta_{nuisance}$) and enable discrete sequence modeling, we project the continuous drift through a

Finite Scalar Quantization (FSQ) bottleneck. We define a uniform scalar grid bounded by L_{max} with a step size s . Utilizing the stop-gradient operator, $\text{sg}(\cdot)$, we formulate the complete FSQ forward pass and straight-through estimator (STE) backward pass as a single operation:

$$\Delta z_q = \Delta z_{raw} + \text{sg} \left(s \cdot \text{clip} \left(\lfloor \frac{\Delta z_{raw}}{s} \rfloor, -L_{max}, L_{max} \right) - \Delta z_{raw} \right). \quad (5)$$

The rounding operation $\lfloor \cdot \rfloor$ acts as a topological dead-zone: any continuous noise fluctuation where $|\Delta z_{raw}| < s/2$ is mapped exactly to zero.

Finally, to reconstruct the future state, the quantized drift is added back to the baseline continuous latent and passed through the de-tokenizer $\mathcal{D}$:

$$\hat{x}_{fut} = \mathcal{D}(z_{cur} + \Delta z_q). \quad (6)$$

This residual objective forces the autoencoder to allocate its discrete code-book capacity to the semantic structural changes (Δz_q), keeping the target separate from the stationary background.

4.2 Autoregressive Drift Generation

With the Latent Drift Tokenizer trained, we extract the discrete drift tokens Δz_q for all longitudinal pairs in the training corpus. The forecasting task then becomes a discrete sequence modeling problem over the isolated biological progression [41].

We use a Decoder-Transformer, $\mathcal{G}_\theta$, to autoregressively synthesize the quantized drift. The model is conditioned on the continuous baseline anatomy z_{cur} and a fused embedding of the clinical metadata c and time horizon Δt :

$$P(\Delta \hat{z}_q \mid x_{cur}, c, \Delta t) = \mathcal{G}_\theta(\Delta \hat{z}_q \mid z_{cur}, \text{MLP}(c, \Delta t)). \quad (7)$$

By optimizing the cross-entropy loss over the FSQ vocabulary, the Transformer models the probability distribution of semantic changes. During inference, the predicted drift tokens are decoded and added to the patient’s baseline scan, yielding a patient-specific forecasted MRI.

4.3 Theoretical Resolution: FSQ as a Topological Dead-Zone Filter

Section 3 established a structural limitation: continuous Lipschitz predictors trained on $\Delta z_{raw} = \delta + \eta$ must interpolate the dense nuisance component η , thereby destroying the sparsity of the true biological drift δ .

We now show that applying Finite Scalar Quantization (FSQ) to the latent residual introduces the precise non-Lipschitz behavior required to break this failure mode.

Dead-Zone Quantization. Let $\delta \in \mathbb{R}^d$ be s -sparse with minimum signal magnitude $\gamma > 0$, and let η be dense nuisance noise satisfying $\|\eta\|_\infty \leq \varepsilon$ with $\varepsilon < \gamma$. The separation condition $\varepsilon < \gamma$ corresponds to a minimal signal strength (β -min) assumption commonly imposed in exact support recovery analyses under noise [57]. Without such a margin condition, sparse support is statistically non-identifiable regardless of the estimator.

Define a scalar quantizer with step size h :

$$Q_h(z)_j = h \cdot \left\lfloor \frac{z_j}{h} \right\rfloor. \quad (8)$$

We calibrate the quantization scale such that

$$h > 2\varepsilon. \quad (9)$$

Theorem 2 (FSQ as a Topological Dead-Zone Filter (Informal, see Appendix for formal version)). *Under the conditions above, quantization of the residual*

$$\Delta z_q = Q_h(\delta + \eta) \quad (10)$$

has the following properties:

1. *All nuisance-only coordinates vanish:*

$$j \notin \text{supp}(\delta) \Rightarrow \Delta z_{q,j} = 0. \quad (11)$$

2. *The support of the biological drift is preserved:*

$$\text{supp}(\Delta z_q) = \text{supp}(\delta). \quad (12)$$

3. *The dense noise manifold collapses to a sparse discrete set prior to learning.*

Consequently, the generative model no longer receives dense stochastic supervision, and the Continuous Interpolation Trap no longer applies.

Intuitive Remark For coordinates outside the true pathology support, the residual equals pure noise with magnitude strictly below $h/2$, forcing quantization to zero. On the true support, the signal magnitude exceeds the dead-zone threshold, ensuring non-zero discrete activation. Since Q_h is discontinuous at the bin boundaries, it is non-Lipschitz, which breaks the smooth interpolation behavior of standard neural predictors.

Implication. FSQ is more than a compression mechanism. Applied to the latent drift, it acts as a structural projection that eliminates dense nuisance noise, restores sparse support geometry, and converts continuous regression into discrete event modeling. By applying this topological dead-zone before autoregressive modeling, Quantized Latent Drift addresses the Continuous Interpolation Trap and supports structurally faithful forecasting.

Table 1: Quantitative comparison of neuro-forecasting performance on the longitudinal 3D brain MRI dataset. We evaluate models across three dimensions: (i) **Generative Fidelity** via FID (i3d and classifier-based), (ii) **Structural Similarity** between forecasted and ground-truth progression, and (iii) **Downstream Clinical Utility** via AD diagnosis performance. Bold and underlined values indicate the best and second-best results, respectively.

Method	Generative Fidelity		Structural Similarity			Downstream Clinical Utility			
	FID (i3d)	FID (cls.)	Diff-SSIM	Pix-SSIM	NCC	F_1	Pre.	Rec.	Acc.
<i>CycleGAN</i> [58]	12.13	7.03	0.7713	0.9612	0.9611	75.44	85.95	67.22	79.39
Diffusion based solutions									
<i>Palette</i> [44]	46.74	9.27	0.8030	0.8088	0.9810	80.36	<u>87.50</u>	74.30	82.89
<i>I2I-DiT</i> [34]	25.65	<u>7.63</u>	0.7799	0.7979	0.9807	<u>62.37</u>	87.00	48.60	72.36
<i>BrLP</i> [39]	19.59	10.14	0.8063	0.8574	<u>0.9819</u>	80.36	86.98	74.67	82.81
Auto-regressive based solutions									
<i>VQGAN</i> [13]	30.11	18.29	0.7625	0.8612	0.9576	70.55	78.23	64.24	74.74
<i>RQ-Transformer</i> [27]	27.71	10.89	0.7071	0.7507	0.8666	60.78	62.82	58.84	64.21
<i>NeuroAR</i> [55]	22.28	15.34	<u>0.8128</u>	0.8554	0.9782	<u>81.86</u>	87.50	<u>76.91</u>	83.95
<i>Ours</i>	<u>13.92</u>	12.77	0.8204	<u>0.9483</u>	0.9880	87.51	88.26	86.78	88.33

5 Experiments

5.1 Evaluation Setup

Benchmarks. We utilize data from the ADNI [21] and AIBL [12] datasets. During preprocessing, all 3D MRI volumes are downsampled to a resolution of $93 \times 112 \times 96$. By leveraging the patients’ longitudinal follow-up information, we construct a total of 3,981 current-future scan pairs. From this cohort, 1,140 pairs are allocated to the testing set.

Evaluation metrics. We evaluate the generated 3D MRIs across three dimensions: Structural Similarity, which assesses spatial agreement between generated and ground truth volumes via SSIM and NCC metrics; Downstream Clinical Utility, which measures the diagnostic value of synthesized images using a pre-trained Alzheimer’s Disease classification model; and Generative Fidelity, which evaluates distribution-level realism using Fréchet Inception Distance (FID) with features extracted by I3D and our domain-specific classifier (see Appendix for details).

Clinical evaluation protocol. For Downstream Clinical Utility, we use a frozen AD classifier (a ViViT-style model with over 91% accuracy on real scans) that is run on each generated future scan, and the predicted label is compared against the patient’s real ground-truth diagnosis at the matching time. The classifier and forecasting cohorts are patient-disjoint, so no patient seen during classifier training appears in the forecasting test set, and we draw a single forecast per (patient, target time) without multi-sample averaging or ensembling.

5.2 Main Results

The results in Table 1 support our analysis of Identity Collapse and the Continuous Interpolation Trap. Latent Drift achieves the best **Diff-SSIM (0.8204)** and **NCC (0.9880)**, which indicates that targeting the latent temporal residual Δz helps the model avoid being dominated by stationary anatomy. CycleGAN [58] achieves higher generative fidelity in terms of **FID (12.13 and 7.03)**, but its lower **Diff-SSIM (0.7713)** suggests that its optimization mainly reproduces static neuroanatomy rather than capturing the sparse biological drift needed for forecasting. Wilcoxon signed-rank tests over patient-level Accuracy, F_1 , and Diff-SSIM confirm that these gains are statistically significant, with our method outperforming every baseline at $p < 0.05$.

The clinical metrics point to the role of the Finite Scalar Quantization (FSQ) step. Acting as a topological dead-zone filter, FSQ suppresses the dense nuisance noise that drives continuous models toward spurious variance, which helps keep the predicted trajectory structurally consistent. Our method leads on F_1 (**87.51**) and **accuracy (88.33)** over diffusion-based baselines such as Palette [44] and I2I-DiT [34]. Overall, separating biological progression from baseline anatomy through a quantized residual is an effective strategy for patient-specific neuro-forecasting.

Longitudinal Trajectory Fidelity Figure 5 shows the longitudinal trajectory analysis, which tests whether the framework can simulate patient-specific neurodegeneration over long clinical intervals. We track the structural similarity between the baseline scan and the forecasted states over a four-year period, and find that our approach follows the rate of anatomical decline with good accuracy and tracks the ground-truth progression closely.

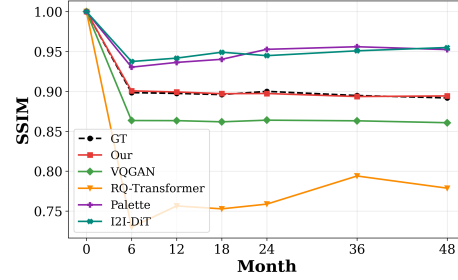


Fig. 5: Personalized Longitudinal Progression Trajectory.

This suggests the model avoids the Identity Collapse seen in competing methods. The high SSIM of the Palette and I2I-DiT trajectories indicates that these models settle into a near-identity mapping, mostly reproducing stationary anatomy instead of the sparse pathological signal δ .

In contrast, the VQGAN [13] and RQ-Transformer [27] baselines drift and fluctuate, showing that standard continuous and autoregressive models struggle to stay anatomically consistent under low-signal conditions. Our framework remains stable over the 48-month horizon, which is consistent with the role of the FSQ bottleneck in preventing the spread of spurious variance across the volume. By recovering the personalized atrophy trajectory, Latent Drift provides the structural consistency needed for clinical forecasting.

Consistency Across Fast- and Slow-Changing Regions A natural concern with a globally uniform FSQ grid is that anatomical regions with very different rates

of change might be quantized inconsistently. Because FSQ is applied in the learned latent space rather than in pixel space, the encoder maps anatomically heterogeneous regions onto a comparable scale before the dead-zone is applied. Fig. 6 reports the recovered-to-ideal change ratio on three regions of interest that span roughly a $23\times$ range of dynamic change (hippocampus, cerebellum, and ventricle). Our method stays within $[0.93, 1.18]$ of the ideal ratio across all three regions, with a mean absolute deviation of 0.12, which is 2.6 to $5.8\times$ closer to the ideal than the competing baselines. This indicates that the shared grid does not systematically under-represent fast-changing regions in practice, though a region-aware quantization scheme remains an open direction.

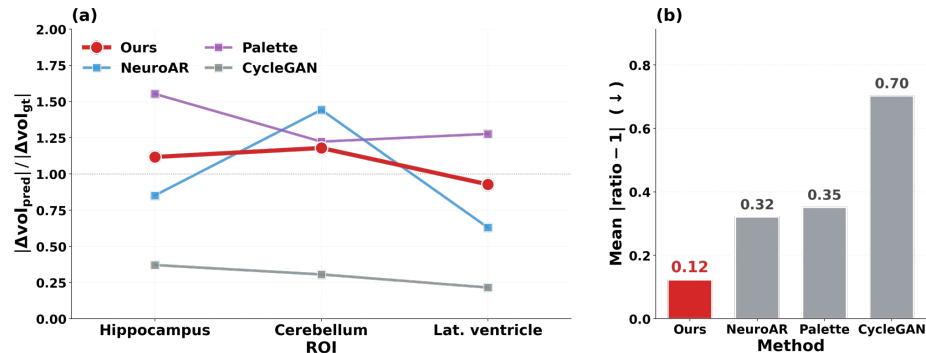


Fig. 6: Recovered-to-ideal change ratio across three regions of interest spanning a wide range of dynamic change. Our method stays close to the ideal ratio of 1 across regions, indicating that the shared latent-space quantization grid handles anatomically heterogeneous regions consistently.

5.3 Ablation Studies

Which prediction target is better? Table 2 supports moving from pixel-level synthesis to latent drift prediction, which addresses the Identity Collapse from our analysis. When the model

Table 2: Ablation of Prediction Targets.

Target	FID	Diff-SSIM	Acc.	F_1
Pixel	44.38	0.7625	77.19	72.75
Drift (ours)	13.92	0.8205	88.33	87.52

performs direct pixel-level forecasting, it reaches a higher FID of 44.38 and a lower Diff-SSIM of 0.7625. This indicates that the optimization is dominated by reproducing stationary neuroanatomy rather than capturing faint biological progression. Shifting the target to the latent temporal residual Δz concentrates model capacity on the progression-relevant dynamics.

How does FSQ perform comparing with other quantization methods? The ablation results in Table 3 support the choice of Finite Scalar Quantization as the topological dead-zone filter for neuro-forecasting. While Vector Quantization (VQ) and Lookup-Free

Quantization (LFQ) achieve slightly higher Diff-SSIM scores of 0.8249 and 0.8253 respectively, FSQ provides the best overall generative fidelity with an rFID of 13.77. Binary Spherical Quantization (BSQ) reaches an rFID of 20.70 and an F_1 of 82.75, also below FSQ. This suggests that while standard VQ variants can capture structural changes, FSQ better preserves the semantic compressed representation required for high-fidelity MRI synthesis.

Table 3: Ablation of Quantization Mechanisms

Quantizers	rFID	Diff-SSIM	Acc.	F_1
FSQ (ours)	13.77	0.8224	88.24	87.41
VectorQuan.	14.04	0.8249	86.67	85.52
Residual VQ	29.06	0.7798	84.56	83.67
LookupFreeQuan.	19.09	0.8253	85.09	84.02
BSQ	20.70	0.7795	84.56	82.75

Table 4: Ablation of FSQ grid configurations across reconstruction (left) and generative forecasting (right). The configuration [8, 8, 8, 5, 5, 5] (highlighted) achieves the optimal trade-off between representational capacity and noise suppression, leading to superior clinical diagnostic performance.

(a) Reconstruction Performance (Stage 1)

FSQ Size	rFID	Diff-SSIM	Acc.	F_1
[8,6,5]	6.93	0.8319	86.49	84.90
[8,5,5,5]	14.95	0.8528	82.89	80.04
[7,5,5,5,5]	6.36	0.8520	86.67	85.19
[8,8,8,6,5]	24.24	0.8579	82.54	79.46
[8,8,8,5,5,5]	9.45	0.8536	87.98	87.01

(b) Generative Forecasting (Stage 2)

FSQ Size	rFID	Diff-SSIM	Acc.	F_1
[8,6,5]	7.07	0.8128	86.84	85.27
[8,5,5,5]	15.58	0.8240	82.46	79.51
[7,5,5,5,5]	7.00	0.8253	85.35	83.64
[8,8,8,6,5]	27.07	0.8273	82.81	80.00
[8,8,8,5,5,5]	12.78	0.8224	88.24	87.41

Which FSQ size achieves the best trade-off between generation and reconstruction? Table 4 shows a trade-off: the FSQ step size must suppress high-frequency noise while preserving sparse biological drift. Coarse configurations (e.g., [8, 6, 5]) discard subtle indicators, while high-capacity grids (e.g., [8, 8, 8, 6, 5]) degrade stability (rFID 27.07) by interpolating imaging noise, which is the Continuous Interpolation Trap.

Configuration [8, 8, 8, 5, 5, 5] balances expressive power with signal isolation through a non-Lipschitz threshold, reaching the best clinical metrics (Accuracy 88.24%, F_1 87.41%). Other configurations yield higher structural similarity but lose diagnostic reliability by capturing non-semantic fluctuations. Calibrated quantization therefore matters for neuro-forecasting as more than a compression step.

Which prediction target is better? Table 5 compares generation paradigms and shows that a high-capacity, conditioned sequence model is needed to forecast the discrete progression signal. While the ContextCon. baseline reaches a competitive rFID of 14.49,

Table 5: Comparison of Generation Paradigms.

Generator	rFID	Diff-SSIM	Acc.	F_1
ContextCon.	14.49	0.7954	83.77	81.92
MaskedCon.	23.17	0.8169	88.07	87.41
Ours	13.92	0.8205	88.33	87.52

its relatively low Diff-SSIM of 0.7954 and Accuracy of 83.77% suggest that simple conditioning on baseline anatomy is insufficient to capture the sparse pathological progression. Conversely, the MaskedCon. variant improves structural capturing, yet it suffers from a significant degradation in generative fidelity. This trade-off reflects the difficulty of maintaining global anatomical consistency while localized pathological changes are being synthesized.

6 Related Work

Static brain MRI generation. The first generation of work established that deep generative models can produce anatomically realistic brain volumes. 3D GANs combining variational autoencoders with adversarial training were shown to synthesize whole-brain MRIs that capture the statistical distribution of healthy and pathological anatomy [26, 59]. Conditional diffusion probabilistic models subsequently improved both fidelity and diversity, generating brain images whose distributional properties closely match those of real clinical scans [37]. In parallel, cross-modality translation networks demonstrated that generative models can learn the highly nonlinear mapping between MRI contrasts [47, 54]. These results show that modern generative architectures have enough representational capacity to model 3D neuroanatomy. However, all of the above operate on single time points: they produce a plausible brain image but encode no notion of how that anatomy evolves over time.

Conditioned brain aging and disease simulation. A second line of work introduces temporal and pathological conditioning to steer generation. Age-conditioned GANs learn to transform an input scan into what the same brain might look like at a different age by disentangling identity-preserving content from age- and sex-dependent style [17]. CounterSynth extends this idea to counterfactual reasoning: a conditional model of diffeomorphic deformations modifies select morphological features of a brain image to reflect a specified demographic or diagnostic label while leaving subject-specific details intact [38]. Diffusion-based counterfactual generation further enables interpretable detection of Alzheimer’s disease effects by comparing an individual’s real scan with a model-generated healthy counterfactual of matched age and sex [9]. Most recently, InBrainSyn combines cohort-level template learning with parallel transport to simulate individualized aging trajectories under both healthy and disease conditions from a single observation [16]. These methods demonstrate that generative models can be meaningfully conditioned on clinically relevant variables.

Individualized longitudinal prediction. TimeFlow conditions a U-Net on continuous age variables under symmetric diffeomorphic constraints, enabling nonlinear extrapolation of individual atrophy trajectories [22]. Deformation-aware diffusion models incorporate explicit morphological priors into the generation process to better capture the characteristic atrophy patterns of Alzheimer’s disease [20]. An alternative strategy bypasses the deformation assumption and directly generates future voxel intensities: conditional GANs with 3D discriminators have been trained to synthesize whole-brain MRIs at a target future time

point [23]. The autoregressive NeuroAR framework tokenizes 3D brain volumes and predicts future token maps via cross-attention on acquisition and target ages, outperforming both GAN and latent-diffusion baselines across multiple longitudinal cohorts [55]. Brain Latent Progression further operates in a compressed latent space to model individualized spatiotemporal disease trajectories via conditional latent diffusion [40].

7 Conclusion

This work identifies and addresses two optimization pathologies, Identity Collapse and the Continuous Interpolation Trap, that cause standard generative models to fail when forecasting slow-evolving neurodegeneration. We introduce Latent Drift, a progressive framework that shifts the predictive target from absolute pixel-level anatomy to a compressed temporal residual. By applying Finite Scalar Quantization as a non-Lipschitz topological dead-zone filter, our approach suppresses dense nuisance noise while preserving the sparse biological progression. Experiments on longitudinal 3D brain MRIs show that Latent Drift improves forecasting fidelity and clinical diagnostic accuracy over diffusion and autoregressive baselines, providing a patient-specific forecasting tool that can support clinical trial design and earlier intervention.

Limitations and Future Work. Our shared dead-zone quantization grid assumes a stable background, so it can under-represent regions that change quickly within a clinical interval, such as the ventricles, while still filtering subtle change elsewhere. We also validate only on well-registered, downsampled ADNI and AIBL cohorts, leaving open how to separate subtle progression from large site-specific deformation (e.g., gradient nonlinearity) in raw multi-site data. Future work will address both through a region-aware quantization grid and validation on raw multi-site scans, and will extend the framework to other slow-evolving pathologies and modalities such as chest X-ray and retinal imaging.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2025YFF0511302). This work was partially supported by RGC Collaborative Research Fund (No. C5055-24G), the Start-up Fund of The Hong Kong Polytechnic University (No. P0045999), the Seed Fund of the Research Institute for Smart Ageing (No. P0050946), and Tsinghua-PolyU Joint Research Initiative Fund (No. P0056509), and PolyU UGC funding (No. P0053716).

References

1. Apostolova, L.G., Green, A.E., Babakchanian, S., Hwang, K.S., Chou, Y.Y., Toga, A.W., Thompson, P.M.: Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (MCI), and Alzheimer disease. *Alzheimer Disease & Associated Disorders* **26**, 17–27 (2012)
2. Assunção, S.S.M., Sperling, R.A., Ritchie, C.W., Kerwin, D.R., Aisen, P.S., Lansdall, C.J., Atri, A., Cummings, J.L.: Meaningful benefits: a framework to assess disease-modifying therapies in preclinical and early Alzheimer’s disease. *Alzheimer’s Research & Therapy* **14** (2022)
3. Bateman, R.J., Xiong, C., Benzinger, T.L., Fagan, A.M., Goate, A.M., Fox, N.C., Marcus, D.S., Cairns, N.J., Xie, X., Blazey, T.M., Holtzman, D.M., Santacruz, A.M., Buckles, V., Oliver, A., Moulder, K.L., Aisen, P.S., Ghetti, B.F., Klunk, W.E., McDade, E.M., Martins, R.N., Masters, C.L., Mayeux, R.P., Ringman, J.M., Rossor, M.N., Schofield, P.R., Sperling, R.A., Salloway, S.P., Morris, J.C.: Clinical and biomarker changes in dominantly inherited Alzheimer’s disease. *The New England journal of medicine* **367** **9**, 795–804 (2012)
4. Blinkouskaya, Y., Caçoilo, A., Gollamudi, T., Jalalian, S., Weickenmeier, J.: Brain aging mechanisms with mechanical manifestations. *Mechanisms of ageing and development* **200**, 111575 – 111575 (2021)
5. Brokman, J., Rachmil, O., Hofman, O., Betser, R., Giloni, A., Vainshtein, R., Kojima, H.: Training-free detection of text-to-video generations via over-coherence. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 3993–4003 (2026)
6. Caldwell, J.C., Miller, M.M., Wing, S., Soll, D.R., Eberl, D.F.: Dynamic analysis of larval locomotion in drosophila chordotonal organ mutants. *Proceedings of the National Academy of Sciences* **100**(26), 16053–16058 (2003)
7. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 10850–10869 (2022)
8. Cui, J., Wu, J., Li, M., Yang, T., Li, X., Wang, R., Bai, A., Ban, Y., Hsieh, C.J.: LoL: Longer than longer, scaling video generation to hour. *arXiv preprint arXiv:2601.16914* (2026)
9. Dhinagar, N.J., Thomopoulos, S.I., Laltoo, E., Thompson, P.M.: Counterfactual MRI generation with denoising diffusion models for interpretable Alzheimer’s disease effect detection. 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) pp. 1–6 (2024)
10. DiStefano III, J.: *Dynamic systems biology modeling and simulation*. Academic Press (2015)
11. Dubois, B., Hampel, H., Feldman, H., Scheltens, P., Aisen, P.S., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavado, E., Crutch, S.J., Dartigues, J.F., Duyckaerts, C., Epelbaum, S., Frisoni, G.B., Gauthier, S., Genthon, R., Gouw, A.A., Habert, M.O., Holtzman, D.M., Kivipelto, M., Lista, S., Molinuevo, J.L., O’Bryant, S.E., Rabinovici, G.D., Rowe, C.C., Salloway, S.P., Schneider, L.S.S., Sperling, R.A., Teichmann, M., Carrillo, M.C., Cummings, J.L., Jack, C.R.: Preclinical Alzheimer’s disease: Definition, natural history, and diagnostic criteria. *Alzheimer’s & Dementia* **12** (2016)
12. Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., et al.: The Australian imaging,

- biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. *International psychogeriatrics* **21**(4), 672–687 (2009)
13. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 12873–12883 (2021)
 14. Farki, A., Moradi, E., Koundal, D., Tohka, J.: Forecasting future anatomies: Longitudinal brain MRI-to-MRI prediction. *ArXiv* (2025)
 15. Fazlyab, M., Robey, A., Hassani, H., Morari, M., Pappas, G.J.: Efficient and accurate estimation of lipschitz constants for deep neural networks. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (2019)
 16. Fu, J., Zheng, Y., Dey, N., Ferreira, D., Moreno, R.: Synthesizing individualized aging brains in health and disease with generative models and parallel transport. *Medical image analysis* **105**, 103669 (2025)
 17. Gadewar, S.P., Ramesh, A., Liu, M., Gari, I.B., Nir, T.M., Thompson, P.M., Jahanshad, N.: Predicting individual brain MRIs at any age using style encoding generative adversarial networks. In: *Symposium on Medical Information Processing and Analysis* (2023)
 18. Haji-Ali, M., Menapace, W., Siarohin, A., Balakrishnan, G., Ordonez, V.: Taming data and transformers for audio generation. *International Journal of Computer Vision* **134**(3), 87 (2026)
 19. Henneman, W.J.P., Sluimer, J.D., Barnes, J., van der Flier, W.M., Sluimer, I.C., Fox, N.C., Scheltens, P., Vrenken, H., Barkhof, F.: Hippocampal atrophy rates in Alzheimer disease. *Neurology* **72**, 1007 – 999 (2009)
 20. Honga, X., Lin, J., Wang, M.: Deformation-aware temporal generation for early prediction of Alzheimers disease. *ArXiv abs/2511.21114* (2025)
 21. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., 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)
 22. Jian, B., Pan, J., Li, Y., Bongratz, F., Li, R., Rueckert, D., Wiestler, B., Wachinger, C.: TimeFlow: Temporal conditioning for longitudinal brain MRI registration and aging analysis (2025)
 23. Jung, E., Luna, M., Park, S.H.: Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression. *Pattern Recognit.* **133**, 109061 (2022)
 24. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacıhaliloglu, I., Merhof, D.: Diffusion models in medical imaging: A comprehensive survey. *Medical image analysis* **88**, 102846 (2022)
 25. KM, M.M., Arun, A., Laskar, Z., Jawahar, C.: PhyEduVideo: A benchmark for evaluating text-to-video models for physics education. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 8690–8699 (2026)
 26. Kwon, G., Han, C., Kim, D.S.: Generation of 3d brain MRI using auto-encoding generative adversarial networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2019)
 27. Lee, D., Kim, C., Kim, S., Cho, M., Han, W.S.: Autoregressive image generation using residual quantization. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 11523–11532 (2022)

28. Leung, K.K., Bartlett, J.W., Barnes, J., Manning, E.N., Ourselin, S., Fox, N.C.: Cerebral atrophy in mild cognitive impairment and Alzheimer disease. *Neurology* **80**, 648 – 654 (2013)
29. Ma, Y., Feng, K., Hu, Z., Wang, X., Wang, Y., Zheng, M., Wang, B., Wang, Q., He, X., Wang, H., et al.: Controllable video generation: A survey. *arXiv preprint arXiv:2507.16869* (2025)
30. Ma, Y., He, Y., Cun, X., Wang, X., Chen, S., Li, X., Chen, Q.: Follow your pose: Pose-guided text-to-video generation using pose-free videos. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 38, pp. 4117–4125 (2024)
31. Ma, Y., Liu, Y., Zhu, Q., Yang, A., Feng, K., Zhang, X., Yan, Z., Li, Z., Han, S., Qi, C., et al.: Follow-your-motion: Video motion transfer via efficient spatial-temporal decoupled finetuning. *arXiv preprint arXiv:2506.05207* (2025)
32. Nakashima, S., Sato, K., Niimi, Y., Toda, T., Iwatsubo, T.: Therapeutic time window of disease-modifying therapy for early Alzheimer’s disease. *Alzheimer’s & Dementia* **21** (2025)
33. Narayanan, S., Nakamura, K., Nakamura, K., Fonov, V.S., Maranzano, J., Caramanos, Z., Giamini, P.S., Collins, D.L., Arnold, D.L.: Brain volume loss in individuals over time: Source of variance and limits of detectability. *NeuroImage* p. 116737 (2020)
34. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 4195–4205 (2023)
35. Peng, W., Adeli, E., Zhao, Q., Pohl, K.M.: Generating realistic brain MRIs via a conditional diffusion probabilistic model. *International Conference on Medical Image Computing and Computer-Assisted Intervention* **14227**, 14–24 (2022)
36. Pereira, J.B., Hall, S., Jalakas, M., Grothe, M.J., Strandberg, O., Stomrud, E., Westman, E., van Westen, D., Hansson, O.: Longitudinal degeneration of the basal forebrain predicts subsequent dementia in Parkinson’s disease. *Neurobiology of disease* p. 104831 (2020)
37. Pinaya, W.H.L., Tudosiu, P.D.D., Dafflon, J., Costa, P.F.D., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain imaging generation with latent diffusion models. *ArXiv abs/2209.07162* (2022)
38. Pombo, G., Gray, R.J., Cardoso, M.J., Ourselin, S., Rees, G.E., Ashburner, J., Nachev, P.: Equitable modelling of brain imaging by counterfactual augmentation with morphologically constrained 3D deep generative models. *Medical Image Analysis* **84** (2021)
39. Puglisi, L., Alexander, D.C., Ravi, D.: Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 173–183. Springer (2024)
40. Puglisi, L., Alexander, D.C., Ravi, D.: Brain latent progression: Individual-based spatiotemporal disease progression on 3D brain mris via latent diffusion. *Medical image analysis* **106**, 103734 – 103734 (2025)
41. Qian, Y., Wang, J., Feng, Y., Xu, C., Lu, W., Liu, Y., Sun, B., Chen, Y., Liu, Y., Wang, S.: Think before you move: Latent motion reasoning for text-to-motion generation. *arXiv preprint arXiv:2512.24100* (2025)
42. Rohrer, J.D.: Structural brain imaging in frontotemporal dementia. *Biochimica et biophysica acta* **1822** **3**, 325–32 (2012)
43. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 10674–10685 (2021)

44. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 conference proceedings. pp. 1–10 (2022)
45. Scaman, K., Virmaux, A.: Lipschitz regularity of deep neural networks: analysis and efficient estimation. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. p. 3839–3848 (2018)
46. Schmidt-Rhaesa, A.: The evolution of organ systems. OUP Oxford (2007)
47. Sharma, A., Hamarneh, G.: Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Transactions on Medical Imaging* **39**, 1170–1183 (2019)
48. de Vugt, M.E., Verhey, F.R.J.: The impact of early dementia diagnosis and intervention on informal caregivers. *Progress in neurobiology* **110**, 54–62 (2013)
49. Wan, T., Wang, A., Ai, B., Wen, B., Mao, C., Xie, C.W., Chen, D., Yu, F., Zhao, H., Yang, J., et al.: Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314 (2025)
50. Wang, J., Hu, Z., Xu, C., Ren, S., Feng, Y., Liu, Y., Sun, B., Wang, S.: Guided by the plan: Enhancing faithful autoregressive text-to-audio generation with guided decoding. arXiv preprint arXiv:2601.14304 (2026)
51. Wang, J., Xu, C., Yu, C., Hu, Z., Xie, H., Yu, G., Shang, L., Wang, S.: Language model based text-to-audio generation: Anti-causally aligned collaborative residual transformers. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. pp. 26036–26054 (2025)
52. Wang, J., Xu, C., Yu, C., Shang, L., Hu, Z., Wang, S., Bo, L.: Synchronized video-to-audio generation via mel quantization-continuum decomposition. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 3111–3120 (2025)
53. Wittens, M.M.J., Allemeersch, G.J., Sima, D.M., Naeyaert, M., Vanderhasselt, T., Vanbinst, A.M., Buls, N., Brucker, Y.D., Raeymaekers, H., Fransen, E.A., Smeets, D., van Hecke, W., Nagels, G., Bjerke, M., de Mey, J., Engelborghs, S.: Inter- and intra-scanner variability of automated brain volumetry on three magnetic resonance imaging systems in Alzheimer’s disease and controls. *Frontiers in Aging Neuroscience* **13** (2021)
54. Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.I.C., Xu, Y.: MRI cross-modality image-to-image translation. *Scientific Reports* **10** (2020)
55. Yesiloglu, R., Peng, W., Islam, M.T., Adeli, E.: Neural autoregressive modeling of brain aging. ArXiv **abs/2507.22954** (2025)
56. Zhang, Y., Chen, J., Li, Y., Jiao, B., Luo, S.: Disease-modifying therapies for Alzheimer’s disease: clinical trial progress and opportunity. *Ageing research reviews* p. 102595 (2024)
57. Zhao, P., Yu, B.: On model selection consistency of LASSO. *The Journal of Machine Learning Research* **7**, 2541–2563 (2006)
58. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
59. Zoghby, M., Erickson, B., Conte, G.: Generative adversarial networks for brain MRI synthesis: Impact of training set size on clinical application. *Journal of Imaging Informatics in Medicine* **37**, 1228 – 1238 (2024)