NeuroLDM-3D: Enhancing Neurological Disease Detection by Leveraging Conditional Latent Diffusion for Brain MRI Synthesis

Jay Sawant[1,2], Erik Kaestner[2], Ezequiel Gleichgerrcht[3], Leonardo Bonilha[4], Carrie R. McDonald[2], and Kyle Hasenstab*[5]

[1]Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA
[2]UC San Diego Health, San Diego, CA, USA
[3]Emory University, Atlanta, GA, USA
[4]University of South Carolina, Columbia, SC, USA
[5]San Diego State University, San Diego, CA, USA

Preprint — Manuscript currently under review.
This teaser contains only the title, abstract, and selected figures.

Abstract

We introduce *NeuroLDM-3D*, a class-conditional latent diffusion framework for synthesizing realistic three-dimensional brain MRIs to improve model training for automated detection of neurological disease. The approach first trains a 3D variational autoencoder (VAE) to compress volumes into a smooth latent space that preserves neuroanatomy and pathological signatures, then learns a transformer-based denoiser (DiT-3D) to generate class-specific latents (Healthy vs. patient), which are decoded into full-resolution scans. Using a multi-site T1-weighted dataset of healthy controls (HC) and patients with temporal lobe epilepsy (TLE), *NeuroLDM-3D* produces anatomically coherent and class-consistent images that retain disease-relevant pathological cues. Compared with adversarial and voxel-space diffusion baselines, the proposed framework achieves higher generative fidelity, reflecting the benefits of latent-space modeling and transformer-based global context. When augmenting training sets, synthetic volumes improve downstream TLE classification performance in limited-data regimes and maintain performance when real data are abundant. Attribution analyses further show that models trained with only synthetic data identify the same medial-temporal and limbic structures associated with TLE, supporting the neurobiological plausibility of the generated images. Overall, these results demonstrate that targeted, class-aware 3D MRI synthesis using latent diffusion can effectively mitigate data scarcity, enhance diagnostic robustness, and enable scalable, anatomically grounded generative modeling for clinical neuroimaging applications.

Note. This is a shortened preprint teaser prepared for web display and does not contain the full manuscript.

Figure 1: Overview of NeuroLDM-3D

*Corresponding author: kahasenstab@sdsu.edu

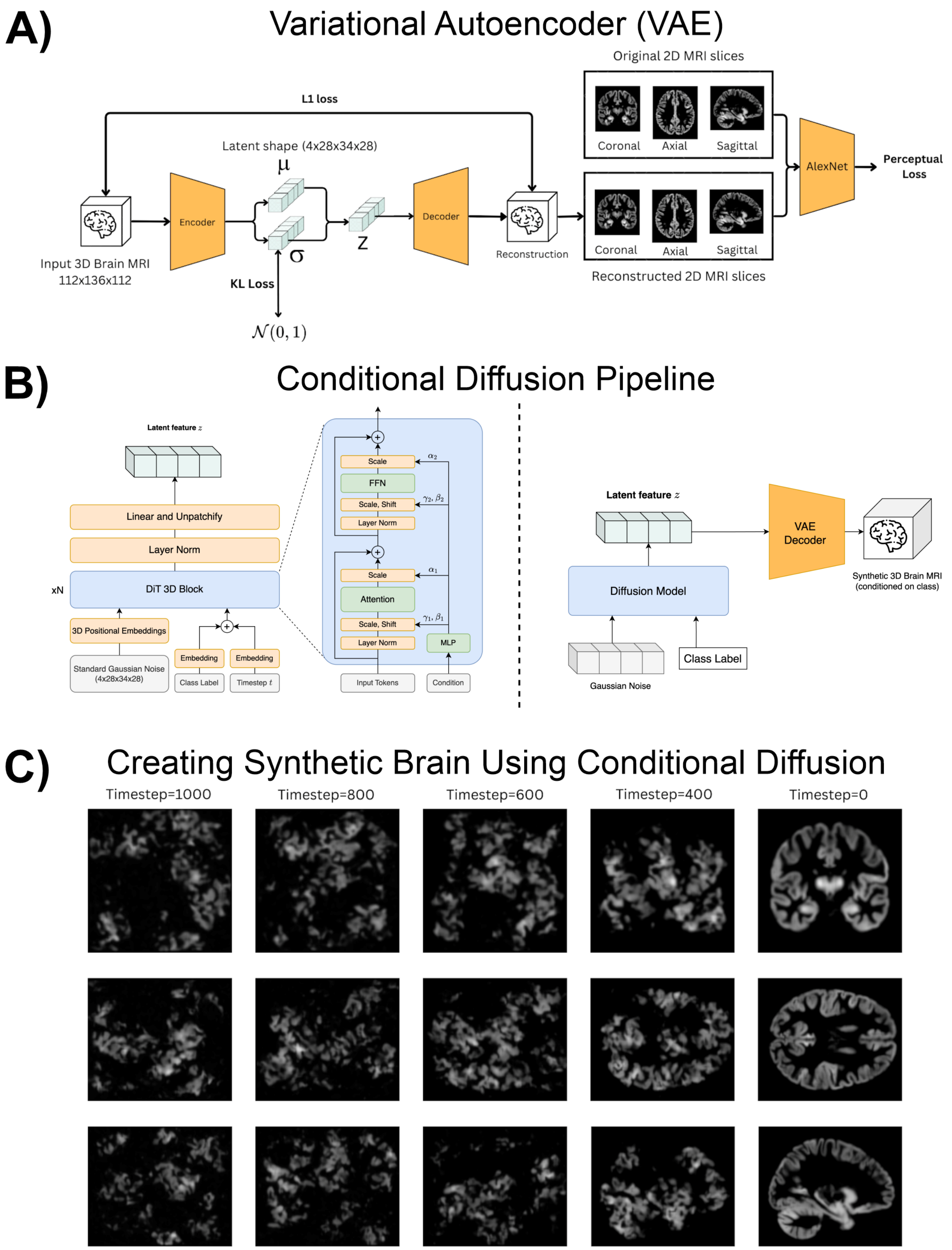

Figure 1: Overview of NeuroLDM-3D. (A) A 3D variational autoencoder (VAE) compresses MRIs into a smooth latent space and reconstructs inputs. **(B)** Class-conditional latent diffusion built on a DiT-3D denoiser: latent tokens are patchified, enriched with 3D positional, timestep, and class embeddings, processed by transformer blocks with Adaptive LayerNorm (AdaLN), and decoded by the VAE to full-resolution scans. **(C)** Denoising trajectory from Gaussian noise (t=1000) to a realistic brain volume (t=0), showing progressive emergence of the structure across the three planes.

Figure 2: Real vs Synthetic 3D Brain MRI Slices

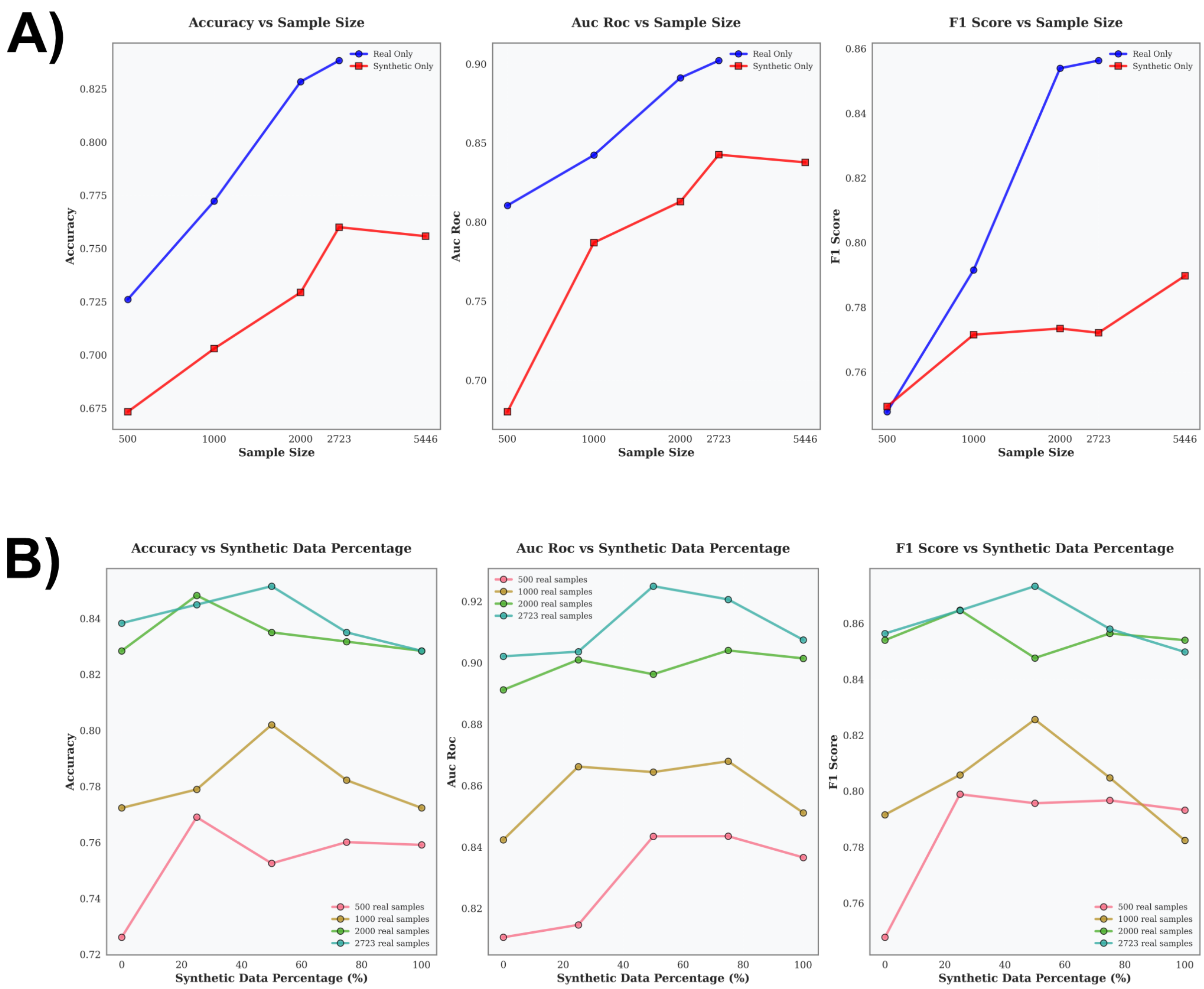

Figure 2: Downstream TLE classification with synthetic augmentation. **(A)** Test Accuracy, AUC–ROC, and F1 as a function of training sample size for two sources: Real only (blue) and Synthetic only (red). Performance with synthetic data steadily improves with scale, indicating high-fidelity, task–useful generation. **(B)** Metrics versus the percentage of synthetic data mixed into the real training set (0, 25, 50, 75, 100%) under four real-data regimes (500, 1000, 2000, 2723 scans). For x real samples, $y\%$ of synthetic data corresponds to $y\%$ of x number of synthetic samples mixed with x real samples. Moderate augmentation (25-50% synthetic) yields the largest gains, especially in low-data settings; whereas heavy replacement ($\geq$75%) tends to plateau or slightly degrade performance. Curves are evaluated on the same held-out test set.