# GraM-Diff: A Unified Graph–Mamba Diffusion Framework for EEG-Based Alzheimer's Disease Data Generation and Diagnosis

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

**Abstract.** Electroencephalography (EEG) is a promising, non-invasive, and cost-effective modality for Alzheimer's disease (AD) detection, but deep learning methods are limited by small and imbalanced clinical datasets. Generative augmentation offers a solution, yet existing approaches rely on inefficient class-specific models or fail to capture complex spatial and temporal brain dynamics. To address this, we propose GraM-Diff, a unified classifier-guided Graph–Mamba diffusion framework for EEG synthesis. It embeds Graph Convolutional Networks within a diffusion U-Net to model inter-electrode connectivity and Bidirectional Mamba state-space blocks for linear-complexity long-range temporal modeling. Latent-space classifier guidance lets a single model generate both healthy and pathological EEG within a shared representation, avoiding fragmented per-cohort pipelines. Across four EEG-based AD benchmarks, synthetic augmentation improves classification, yields superior Context-FID and correlation scores over strong generative baselines, and enhances robustness in data-scarce settings.

**Keywords:** Alzheimer's disease · EEG · diffusion models · generative AI · graph neural networks · state–space models

## 1 Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia worldwide, marked by amyloid-beta plaques and hyperphosphorylated tau tangles that drive synaptic dysfunction [16]. Roughly 7.2 million Americans aged 65+ live with AD, projected to reach 13.8 million by 2060 [3]. Because AD has a long preclinical phase and approved disease-modifying therapies are most effective early, scalable early-diagnostic strategies are critical. Electroencephalography (EEG) is a non-invasive, cost-effective modality for detecting cognitive-decline biomarkers, with accessibility and high temporal resolution well-suited to longitudinal assessment. Computational EEG analysis has progressed from handcrafted spectral/complexity features with classical classifiers [16] to deep CNN [1], GNN [9], and Transformer [20] architectures; Denoising Diffusion Probabilistic Models (DDPMs) have recently emerged for EEG synthesis under limited data [18].

Despite this progress, annotated EEG remains scarce. Existing generative approaches often use *separate* models for healthy and pathological cohorts, increasing overhead and reducing flexibility. Public AD-EEG corpora are small: even combining AD-Auditory, ADSZ, APAVA, and BrainLat yields limited subject counts for high-capacity training [19], while class imbalance and privacy-driven data silos (GDPR, HIPAA) further hinder generalization [17].

To address these challenges, we propose GraM-Diff, a unified Graph–Mamba diffusion framework that combines topology-aware spatial modeling with long-range temporal modeling in a single generative architecture. Unlike perturbation-based augmentation, it directly generates multichannel EEG conditioned on diagnostic state while preserving spatio-temporal structure relevant to downstream classification [15,5]. Our key contributions are: **(i)** a *unified classifier-guided DDPM* that conditions a single generator on diagnostic state via latent-space guidance, enabling scalable class-conditioned multichannel EEG generation within a shared representation rather than per-cohort generators; **(ii)** *topology-aware spatial modeling*, integrating graph-based spatial priors directly into a diffusion U-Net to capture non-Euclidean inter-electrode relationships [9]; **(iii)** *linear-complexity temporal modeling* via Bidirectional Mamba state-space blocks with $\mathcal{O}(L)$ cost over extended recordings [8]; and **(iv)** *noise-robust LogRoPE guidance*, using patch-embedding conditioning with Logarithmic Rotary Positional Embeddings for stable representation learning under diffusion noise [20].

## 2 Related Work

**Discriminative EEG-AD models.** AD diagnosis has evolved from statistical modeling to deep architectures [16], including 3D CNNs on spatio-temporal-spectral tensors [1] and ensembles over raw EEG. These typically treat channels as independent sequences or grid images, ignoring the explicit graph topology of brain connectivity [9]. Multi-granularity Transformers model multiple time scales for EEG-AD assessment [20] but incur quadratic complexity on long recordings and have not been used generatively.

**Data scarcity and generative augmentation.** Aggregated public AD-EEG corpora contain few subjects (often $< 330$), making high-capacity training prone to overfitting [19]. GAN-based augmentation can improve accuracy [5] but suffers training instability and mode collapse and usually synthesizes features rather than raw multichannel signals. Diffusion models offer stabler training and avoid mode collapse [18], and have been applied to EEG super-resolution and psychiatric classification [15]. Yet existing EEG diffusion models are generally not conditioned on AD pathology and lack classifier guidance to retain disease-specific signatures.

Recurrent neural networks and Transformers often struggle with long-sequence EEG modelling, either due to vanishing-gradient issues or the quadratic computational cost of self-attention. Mamba state-space models address these limitations through linear-time sequence modelling, making them suitable for efficient long-range temporal representation learning [8]. In GraM-Diff, we incor-

porate a bidirectional Mamba block within the diffusion bottleneck to capture both causal and anti-causal dependencies across extended EEG recordings. Overall, GraM-Diff combines topology-aware spatial priors, linear-time bidirectional Mamba modelling, and LogRoPE-based multi-granular conditioning within a unified generative–discriminative framework.

## 3 Proposed Methodology

As shown in Fig. 1, GraM-Diff combines spatial electrode relationships and long-range temporal modeling within a single architecture, together with a latent-space guidance mechanism, to generate Alzheimer's stage-conditioned EEG.

### 3.1 Problem Formulation and Diffusion Process

Let the clean multichannel EEG signal be $\mathbf{X}_0 = [x_1, \ldots, x_L] \in \mathbb{R}^{B\times C\times L}$, where $B$ is the batch size, $C$ the number of channels, and $L$ the temporal length. Following the DDPM paradigm to model $q(\mathbf{X}_0)$, the denoising network operates on latent features $\mathbf{h}^t \in \mathbb{R}^{B\times C\times L\times D}$ extracted by the U-Net encoder at timestep $t$. The forward process corrupts $\mathbf{X}_0$ into $\mathbf{X}_t$ over $t \in \{1, \ldots, T\}$ ($T$=1000) as a Gaussian Markov chain:

$$q(\mathbf{X}_t \mid \mathbf{X}_0) = \mathcal{N}\big(\mathbf{X}_t; \sqrt{\bar{\alpha}_t}\,\mathbf{X}_0, (1-\bar{\alpha}_t)\mathbf{I}\big)\,, \qquad \bar{\alpha}_t = \textstyle\prod_{i=1}^{t}(1-\beta_i), \tag{1}$$

with $\beta_i \in (0, 1)$. Because linear schedules abruptly destroy fine-grained high-frequency EEG structure, we adopt a cosine schedule $\bar{\alpha}_t = f(t)/f(0)$, $f(t) = \cos^2\big(\frac{t/T+s}{1+s} \cdot \frac{\pi}{2}\big)$ with $s$=0.008. The network is trained to predict the clean signal $\hat{\mathbf{X}}_0$ rather than the injected noise; the noise term is recovered analytically as $\epsilon_\theta(\mathbf{h}^t, t) = (\mathbf{X}_t - \sqrt{\bar{\alpha}_t}\hat{\mathbf{X}}_0)/\sqrt{1-\bar{\alpha}_t}$ and used in classifier-guided sampling. Diffusion and guidance operate on sequence-first tensors $\mathbb{R}^{B\times L\times C}$, internally permuted to channel-first $\mathbb{R}^{B\times C\times L}$ for U-Net and GCN processing.

### 3.2 Stage 1: Topology-Aware Spatial Encoding (GCN)

Standard convolutions treat EEG channels as grid-adjacent rows, ignoring scalp geometry. We instead pass $\mathbf{X}_t$ through a GCN with electrodes as nodes. Using 3D coordinates $\mathbf{c}_i \in \mathbb{R}^3$ from the International 10–20 system, a geometry-informed adjacency $\mathbf{A} \in \mathbb{R}^{C\times C}$ is built with a distance-based Gaussian kernel,

$$A_{ij} = \begin{cases} \exp\big(-\|\mathbf{c}_i - \mathbf{c}_j\|^2/2\gamma^2\big) & \text{if } \|\mathbf{c}_i - \mathbf{c}_j\| \le \tau \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

with kernel width $\gamma$=2.0 cm and sparsity threshold $\tau$=8.0 cm. We apply the renormalized propagation operator $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A}+\mathbf{I}_C)\tilde{\mathbf{D}}^{-\frac{1}{2}}$, where $\tilde{\mathbf{D}}_{ii} = \sum_j(\mathbf{A}+\mathbf{I}_C)_{ij}$ and the self-loops $\mathbf{I}_C$ preserve each electrode's intrinsic signal. Spatial

aggregation yields $\mathbf{H}_{\text{spatial}} = \text{LayerNorm}(\sigma(\hat{\mathbf{A}}\mathbf{X}_t\mathbf{W}_{\text{gnn}})) \in \mathbb{R}^{B\times C\times D}$ with learnable $\mathbf{W}_{\text{gnn}} \in \mathbb{R}^{L\times D}$ and ReLU $\sigma(\cdot)$. In practice we adopt a fixed, fully connected adjacency as a neutral prior, preserving permutation-invariant message passing while letting the network learn functional connectivity in a data-driven manner; this also admits anatomically informed topologies without architectural change. $\mathbf{H}_{\text{spatial}}$ serves as a spatial inductive bias for temporal modeling.

### 3.3 Stage 2: Hierarchical Temporal Encoder–Decoder

The U-Net encoder is a recurrence over $N$=4 downsampling levels, initialized by $\mathbf{H}^{(0)} = \mathbf{H}_{\text{spatial}}$. The diffusion timestep is encoded with sinusoidal embeddings $\mathbf{t}_{emb}$ projected by a SiLU-MLP and injected into each ResNet block as a learned bias shift, giving $\mathbf{S}^{(l)} = \mathbf{H}^{(l)} + \text{Conv1D}_{out}(\text{SiLU}(\text{GroupNorm}(\mathbf{h}^{(l)}_{res} + \mathbf{h}^{(l)}_{scale})))$, with $\mathbf{h}^{(l)}_{res} = \text{SiLU}(\text{GroupNorm}(\text{Conv1D}_{in}(\mathbf{H}^{(l)})))$ and $\mathbf{h}^{(l)}_{scale} = \text{MLP}(\text{SiLU}(\mathbf{t}_{emb}))$. Each $\mathbf{S}^{(l)}$ is stored as a skip connection and downsampled ($K$=4, stride 2); the compressed code $\mathbf{Z} = \mathbf{H}^{(N)}$ enters the bottleneck. The decoder mirrors the encoder: at each level the map is upsampled by transposed convolution, fused with the matching skip $\mathbf{D}^{(l)}_{fused} = \text{Concat}(\mathbf{D}^{(l)}_{\uparrow}, \mathbf{S}^{(N-1-l)})$, and processed by a time-conditioned ResNet block, starting from $\mathbf{D}^{(0)} = \mathbf{Z}_{out}$ (Stage 3). The decomposition heads then map the output to $\hat{\mathbf{X}}_0 = \mathbf{V}_{trend} + \mathbf{S}_{season}$.

### 3.4 Stage 3: Mamba Bottleneck for Long-Range Temporal Modeling

At the bottleneck, self-attention is replaced by a Bidirectional Mamba block, giving input-dependent long-range modeling with linear complexity $\mathcal{O}(L)$ versus the $\mathcal{O}(L^2)$ cost of Transformers. The underlying SSM has a diagonal state matrix $\mathbf{A} \in \mathbb{R}^{D\times N}$, input matrix $\mathbf{B} \in \mathbb{R}^{D\times N}$, and learns an input-dependent step size $\boldsymbol{\Delta}_k$ per position $k$:

$$\boldsymbol{\Delta}_k = \text{Softplus}(\text{Linear}(\mathbf{Z}_k)), \tag{3}$$

$$\overline{\mathbf{A}}_k = \exp(\boldsymbol{\Delta}_k \odot \mathbf{A}), \quad \overline{\mathbf{B}}_k = (\boldsymbol{\Delta}_k \odot \mathbf{A})^{-1}(\overline{\mathbf{A}}_k - \mathbf{I}) \odot (\boldsymbol{\Delta}_k \odot \mathbf{B}), \tag{4}$$

where $\odot$ is broadcast element-wise multiplication. The adaptive $\boldsymbol{\Delta}_k$ acts as a learned sampling rate: small values capture high-frequency transients, large values model slow background rhythms. We aggregate non-causal context with bidirectional scans $\mathbf{Z}_{mamba} = \text{Linear}(\text{Concat}(\mathbf{h}^{\rightarrow}(\mathbf{Z}), \mathbf{h}^{\leftarrow}(\mathbf{Z})))$ and decompose the latent into a polynomial trend $\mathbf{V}_{trend} = \mathcal{C} \cdot \text{Conv1D}(\text{GELU}(\text{Conv1D}(\mathbf{Z}_{mamba})))$ and a Fourier seasonality $\mathbf{S}_{season} = \mathcal{F}^{-1}(\text{TopK}(\mathcal{F}(\mathbf{Z}_{mamba})))$, with $\mathbf{Z}_{out} = \mathbf{V}_{trend} + \mathbf{S}_{season}$. This separates slow morphological drifts from fast oscillations as an inductive bias, but is trained end-to-end without strict stationarity or periodicity constraints.

### 3.5 Stage 4: Latent-Space Guidance Classifier

Guidance operates on the intermediate latent $\mathbf{h}^t$ rather than the heavily corrupted raw signal $\mathbf{X}_t$, keeping the classifier robust to early diffusion noise while

enforcing pathology-relevant patterns. Given the noisy latent $\mathbf{X}_t \in \mathbb{R}^{B\times L\times C}$, we build multi-granular temporal tokens by partitioning the signal into overlapping patches $\{P_k\}_{k=1}^{G}$ ($G$=3, $P_1$=16, $P_2$=32, $P_3$=64) and projecting each into a shared $D$-dimensional space, $\mathbf{E}_t^{(k)} \in \mathbb{R}^{B\times N_k\times D}$, alongside channel-wise tokens $\mathbf{E}_c^{(m)} \in \mathbb{R}^{B\times C\times D}$ ($M$=1). To stabilize attention over long sequences we apply logarithmically-scaled rotary positional encoding with rotation angle

$$\Theta_j(\text{pos}) = \text{base}^{-\frac{2\lfloor j/2\rfloor}{D}} \cdot \log(\text{pos}+1), \qquad \text{base} = 10000, \tag{5}$$

which moderates positional growth and improves numerical stability. The tokens are concatenated into $\mathbf{H}_0 = \text{Concat}(\mathbf{E}_t^{(1)}, \ldots, \mathbf{E}_t^{(G)}, \mathbf{E}_c^{(1)}, \ldots, \mathbf{E}_c^{(M)}) \in \mathbb{R}^{B\times N\times D}$ ($N = \sum_k N_k + M \cdot C$) and processed by $L_{\text{enc}}$=6 time-conditioned Transformer layers, with $t$ injected through adaptive layer normalization; a global attention-pooling head yields $\mathbf{h} \in \mathbb{R}^{B\times K}$ over $K$ disease categories. During sampling, the noise prediction is modified as

$$\tilde{\epsilon}_t = \epsilon_\theta(\mathbf{h}^t, t) - s\sqrt{1-\bar{\alpha}_t}\,\nabla_{\mathbf{h}^t} \log p_\phi(y \mid \mathbf{h}^t), \tag{6}$$

where $s \in [1, 5]$ controls guidance strength and the factor $\sqrt{1-\bar{\alpha}_t}$ preserves stability. To preserve oscillatory structure we optimize a dual-domain loss combining time-domain reconstruction and frequency-domain spectral constraints,

$$\mathcal{L} = \|\mathbf{X}_0 - \hat{\mathbf{X}}_0\|_1 + \lambda \sum_{p\in\{\text{Re},\text{Im}\}} \|p(\mathcal{F}(\mathbf{X}_0)) - p(\mathcal{F}(\hat{\mathbf{X}}_0))\|_1, \tag{7}$$

with $\mathcal{F}$ the discrete Fourier transform and $\lambda$=0.1. Predicting $\mathbf{X}_0$ directly improves training stability and enables spectral preservation for physiologically realistic synthesis.

## 4 Experiments and Results

### 4.1 Datasets and Implementation

We evaluate on three public EEG datasets, all on a single NVIDIA RTX A4500 GPU (40 GB). **ADSZ** [2]: 19-channel EEG at 128 Hz, AD binary subset of 48 subjects (24 AD / 24 controls), 8–14 s trials. **APAVA** [7]: 16-channel EEG at 256 Hz, 23 subjects (12 AD / 11 controls), 5 s trials. **ADFD** [13]: 19-channel, 500 Hz, 88 subjects (36 AD, 23 FTD, 29 controls), long trials band-passed to 0.5–45 Hz; excluding FTD yields **ADFD-Binary** (65 subjects, 53,182 samples). The generator uses Adam ($\beta_1$=0.9, $\beta_2$=0.96) over 25,000 epochs with EMA weights; the downstream classifier is trained separately for 3,000 epochs (Adam, lr $1.0 \times 10^{-4}$).

### 4.2 Reliability of Generated EEG for AD Diagnosis

We use a subject-independent 80–20 split, train the diffusion model on the 80% subset, and synthesize ∼1.5× extra samples. Classifiers are trained *only* on synthetic data and evaluated on the unseen 20% real test set, comparing "Synthetic

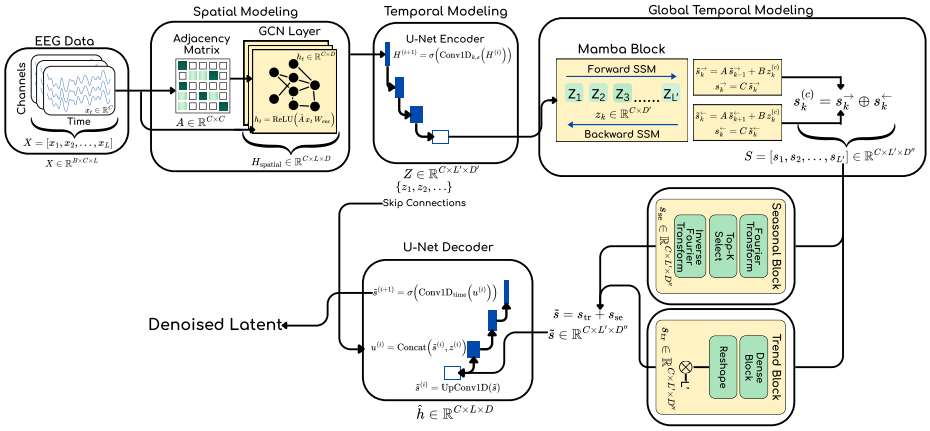


Fig. 1: Architecture of GraM-Diff. EEG data are first passed through a GCN to obtain spatially enriched features. The U-Net encoder extracts hierarchical temporal representations, which enter a bidirectional Mamba bottleneck. The decoder reconstructs the denoised latent and decomposes it into trend and seasonal components. A time-conditioned transformer classifier operating on latent features provides gradients to guide the diffusion process.

Train (TSTR)" against "Real Train (TRTR)" on identical held-out data (Table 1). Models trained purely on synthetic AD data match or exceed the real-data baseline: ADFormer attains 97.14% on ADSZ under synthetic training, surpassing its real-data result (93.08%) and all competing architectures, and reaches 51.22% on the harder ADFD 3-Class task—indicating the generated signals retain diagnostically meaningful patterns that generalize to unseen subjects.

### 4.3 Comparison with Generative Peer Models

We compare synthesis quality against TimeGAN [23], TimeVAE [6], and Diffusion-TS [24], holding the downstream classifier strictly constant so differences reflect generative fidelity alone (Table 2). As TimeGAN and TimeVAE are single-class generators, we train one model per class and merge outputs for multi-class data. GraM-Diff outperforms all baselines across Accuracy, Precision, F1, AUROC, and AUPRC. On ADSZ it reaches 97.18% accuracy, far above Diffusion-TS (87.41%) under identical conditions and the segregated TimeGAN (69.81%) and TimeVAE (64.25%).

### 4.4 Generative Fidelity

We assess fidelity with Context-FID (distributional distance in deep feature space) and Correlation Score (preservation of spatio-temporal dependencies),

Table 1: Classification accuracy (%) trained on our generated data (Synthetic, TSTR) versus the original dataset (Real, TRTR), tested on identical held-out real data. A representative set of baselines is shown.

| | **ADSZ** | | **APAVA** | | **ADFD Binary** | | **ADFD 3 Class** | |
|---|---|---|---|---|---|---|---|---|
| **Model** | **Syn** | **Real** | **Syn** | **Real** | **Syn** | **Real** | **Syn** | **Real** |
| ADFormer [20] | **97.14** | 93.08 | **82.81** | 80.00 | **73.74** | 75.80 | **51.22** | 55.21 |
| Autoformer [22] | 93.61 | 91.43 | 69.89 | 68.64 | 61.31 | 59.81 | 44.28 | 45.25 |
| Informer [25] | 93.18 | 90.22 | 74.86 | 73.11 | 59.95 | 61.88 | 46.90 | 48.45 |
| ITransformer [12] | 78.91 | 78.79 | 75.01 | 74.55 | 71.44 | 73.85 | 49.97 | 52.60 |
| PatchTST [14] | 79.17 | 79.12 | 69.86 | 67.03 | 60.29 | 59.10 | 44.29 | 44.37 |
| Pyraformer [11] | 91.89 | 90.77 | 81.20 | 79.54 | 65.10 | 66.00 | 49.27 | 51.52 |
| TimesNet [21] | 92.59 | 92.75 | 76.31 | 76.30 | 60.27 | 61.79 | 47.14 | 48.81 |
| EEGNet [10] | 89.24 | 86.81 | 71.22 | 71.81 | 65.28 | 71.91 | 47.26 | 51.78 |
| TCN [4] | 82.18 | 88.90 | 80.29 | 84.71 | 59.21 | 68.71 | 42.38 | 50.17 |

Table 2: Comparison with state-of-the-art generative peer models across ADSZ, APAVA, and ADFD (Binary).

| **Model** | **ADSZ** | | | | | **APAVA** | | | | | **ADFD (Binary)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Acc** | **Prec** | **F1** | **AUROC** | **AUPRC** | **Acc** | **Prec** | **F1** | **AUROC** | **AUPRC** | **Acc** | **Prec** | **F1** | **AUROC** | **AUPRC** |
| TimeGAN [23] | 69.81 | 70.28 | 71.93 | 72.29 | 69.35 | 54.28 | 58.19 | 58.92 | 59.70 | 59.00 | 51.35 | 52.63 | 52.74 | 51.70 | 54.36 |
| TimeVAE [6] | 64.25 | 68.11 | 69.74 | 65.29 | 67.81 | 53.91 | 52.14 | 52.83 | 58.51 | 55.92 | 52.10 | 51.88 | 53.83 | 58.14 | 59.53 |
| Diffusion-TS [24] | 87.41 | 89.25 | 88.65 | 90.83 | 90.24 | 63.87 | 64.99 | 63.81 | 70.96 | 72.39 | 55.39 | 58.31 | 57.12 | 60.25 | 65.41 |
| **Ours** | **97.18** | **97.20** | **96.28** | **97.82** | **95.20** | **82.81** | **85.99** | **84.21** | **89.85** | **88.46** | **73.74** | **75.93** | **77.31** | **77.28** | **79.21** |

lower being better, against Diffusion-TS. GraM-Diff attains markedly lower Context-FID on every dataset (ADSZ 0.109 vs. 0.585; APAVA 0.286 vs. 0.426; ADFD-Binary 0.484 vs. 1.030; ADFD 3-Class 0.529 vs. 0.920) and tighter Correlation Scores (ADSZ 0.275 vs. 0.492; APAVA 0.402 vs. 0.829; ADFD-Binary 0.532 vs. 2.890; ADFD 3-Class 0.583 vs. 2.120), confirming closer agreement with real EEG dynamics.

### 4.5 Downstream Utility, Ablation, and Sensitivity

**Utility and augmentation.** Beyond statistical similarity, we test imputation (masking 10–90%), forecasting (horizons 4–32), and real/synthetic mixture classification. Errors stay low under heavy difficulty: on ADSZ, imputation MSE rises only to 0.108 at 90% masking and forecasting MSE remains 0.0501 at horizon 32. Augmentation accuracy is preserved even in heavily synthetic regimes (96.87% on ADSZ at 80% synthetic), confirming the synthetic data's value for limited clinical datasets.

**Ablation.** Removing components (Table 3, top) degrades performance, with Patch Embeddings + LogRoPE most critical: accuracy drops from 97.18% to 91.81% on ADSZ and from 82.81% to 69.65% on APAVA, underscoring its role in long-range temporal and positional modeling. The GNN and Mamba modules each contribute additional gains.

Table 3: Ablation (top) and Gaussian-noise sensitivity (bottom) on ADSZ and APAVA.

| **Setup** | **ADSZ** | | | **APAVA** | |
|---|---|---|---|---|---|
| | **Acc** | **F1** | **AUROC** | **Acc** | **AUROC** |
| Full Model | **97.18** | **96.28** | **97.82** | **82.81** | **89.85** |
| w/o GNN | 95.80 | 94.99 | 96.82 | 79.31 | 84.83 |
| w/o Mamba | 95.23 | 95.28 | 95.27 | 78.45 | 82.36 |
| w/o Patch+LogRoPE | 91.81 | 92.84 | 95.43 | 69.65 | 67.52 |
| *Gaussian noise sensitivity* | | | | | |
| 0% (Base) | **97.18** | **96.28** | **97.82** | **82.81** | **89.85** |
| 10% | 79.20 | 81.54 | 85.35 | 71.77 | 78.42 |
| 20% | 72.98 | 77.31 | 75.30 | 62.85 | 71.20 |

**Sensitivity.** Injecting Gaussian noise (Table 3, bottom) degrades performance gracefully: on ADSZ the model retains an AUROC of 85.35% under 10% noise (vs. 97.82% baseline), with a similar trend on APAVA, indicating robustness to moderate signal perturbation.

## 5 Conclusion

We presented GraM-Diff, a unified classifier-guided Graph–Mamba diffusion framework coupling topology-aware spatial priors with linear-complexity long-range temporal modeling for conditional multichannel EEG synthesis. A single model generates both healthy and pathological signals within a shared latent space, producing high-fidelity samples that better match real distributions and improve downstream AD classification, with superior Context-FID and correlation scores and graceful robustness under noise. Future work will extend the framework to multimodal and irregularly sampled time series while improving scalability and controllability for clinical deployment.

## References

1. Alghamdi, A.M., Ashraf, M.U., Bahaddad, A.A., Almarhabi, K.A., Al Shehri, W.A., Daraz, A.: A novel approach hybrid of ensemble learning and 3-D CNN mechanism: early-stage diagnosis of Alzheimer's disease using EEG signals. Scientific Reports **15**(1), 35893 (2025)
2. Alves, C.L., Pineda, A.M., Roster, K., Thielemann, C., Rodrigues, F.A.: EEG functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia. Journal of Physics: complexity **3**(2), 025001 (2022)
3. Alzheimer's Association: 2025 alzheimer's disease facts and figures. Alzheimer's & Dementia **21**(4), e70235 (2025). https://doi.org/10.1002/alz.70235, https://doi.org/10.1002/alz.70235, accessed: February 21, 2026

4. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)
5. Carrle, F.P., Hollenbenders, Y., Reichenbach, A.: Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depressive disorder. Frontiers in Neuroscience **17**, 1219133 (2023)
6. Desai, A., Freeman, C., Wang, Z., Beaver, I.: Timevae: A variational auto-encoder for multivariate time series generation. arXiv preprint arXiv:2111.08095 (2021)
7. Escudero, J., Abásolo, D., Hornero, R., Espino, P., López, M.: Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. Physiological measurement **27**(11), 1091–1106 (2006)
8. Gui, Y., Chen, M., Su, Y., Luo, G., Yang, Y.: EEGMamba: Bidirectional state space model with mixture of experts for EEG multi-task classification. arXiv preprint arXiv:2407.20254 (2024)
9. Klepl, D., He, F., Wu, M., Blackburn, D.J., Sarrigiannis, P.: EEG-based graph neural network classification of Alzheimer's disease: An empirical evaluation of functional connectivity methods. IEEE Transactions on Neural Systems and Rehabilitation Engineering **30**, 2651–2660 (2022)
10. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering **15**(5), 056013 (2018)
11. Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A.X., Dustdar, S.: Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In: International conference on learning representations (2021)
12. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625 (2023)
13. Miltiadous, A., Tzimourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D.G., Angelidis, P., Tsipouras, M.G., Glavas, E., Giannakeas, N., et al.: A dataset of scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. Data **8**(6), 95 (2023)
14. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730 (2022)
15. Sharma, G., Dhall, A., Subramanian, R.: Medic: Mitigating eeg data scarcity via class-conditioned diffusion model. In: Deep Generative Models for Health Workshop NeurIPS 2023 (2023)
16. Tanveer, M., Sajid, M., Akhtar, M., Quadir, A., Goel, T., Aimen, A., Mitra, S., Zhang, Y.D., Lin, C.T., Del Ser, J.: Fuzzy deep learning for the diagnosis of Alzheimer's disease: Approaches and challenges. IEEE Transactions on Fuzzy Systems **32**(10), 5477–5492 (2024)
17. Umair, M., Khan, M.S., Hanif, M., Ghaban, W., Nafea, I., Qasem, S.N., Saeed, F.: Privacy–preserving dementia classification from EEG via hybrid–fusion EEGNetv4 and federated learning. Frontiers in Computational Neuroscience **19**, 1617883 (2025)
18. Vetter, J., Macke, J.H., Gao, R.: Generating realistic neurophysiological time series with denoising diffusion probabilistic models. Patterns **5**(9) (2024)
19. Wang, Y., Huang, N., Mammone, N., Cecchi, M., Zhang, X.: LEAD: Large Foundation Model for EEG-Based Alzheimer's Disease Detection. arXiv preprint arXiv:2502.01678 (2025)

20. Wang, Y., Mammone, N., Petrovsky, D., Tzallas, A.T., Morabito, F.C., Zhang, X.: ADformer: A multi-granularity transformer for EEG-based Alzheimer's disease assessment. arXiv preprint arXiv:2409.00032 (2024)
21. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)
22. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems **34**, 22419–22430 (2021)
23. Yoon, J., Jarrett, D., Van der Schaar, M.: Time-series generative adversarial networks. Advances in neural information processing systems **32** (2019)
24. Yuan, X., Qiao, Y.: Diffusion-TS: Interpretable diffusion for general time series generation. arXiv preprint arXiv:2403.01742 (2024)
25. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 11106–11115 (2021)