

Separable Yet Inseparable: An EEG-oriented Head for Transformer-based Time-Series Generative Adversarial Networks

Giorgio Scavalli Veccia

`scavalliveccia@di.uniroma1.it`

Sapienza University of Rome

Alessio Fagioli

Saint Camillus International University of Health and Medical Sciences

Marco Cascio

Link Campus University

Marco Raoul Marini

Sapienza University of Rome

Gianluca Foresti

University of Udine

Research Article

Keywords: Generative Adversarial Networks, EEG, Synthetic Data, Evaluation Metrics

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11269400/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Separable Yet Inseparable: An EEG-oriented Head for Transformer-based Time-Series Generative Adversarial Networks

Giorgio Scavalli Veccia^{1*}, Alessio Fagioli², Marco Cascio³, Marco Raoul Marini¹,
Gianluca Foresti⁴

^{1*}Department of Computer Science, University of Rome 'La Sapienza', Via di Salaria,
113, Rome, 00198, Italy.

¹Department of Medicine, UniCamillus-Saint Camillus International University of Health
Sciences, Via di Sant'Alessandro, 8, Rome, 00131, Italy.

³Department of Human Sciences, Link Campus University, Via del Casale di San Pio V,
44, Rome, 00165, Italy.

⁴Department of Mathematics, Computer Science and Physics, University of Udine, Via
delle Scienze, 20, Udine, 33100, Italy.

*Corresponding author(s). E-mail(s): scavalliveccia@di.uniroma1.it;

Abstract

Synthetic electroencephalography (EEG) data can address privacy, data scarcity, and class imbalance. Current generators rarely target EEG specifically and are evaluated with metrics that ignore its physiology. To address these limitations, we propose EEGGAN-SepH, which extends EEGGAN, a transformer-based Time Series Generative Adversarial Network, with a Separable Output Head (SepH). SepH is a lightweight module built around a separable convolution followed by a learnable temporal smoothing filter, designed to model local temporal dynamics and suppress high-frequency noise. We further introduce an evaluation pipeline that computes the Fréchet Distance (FD) and the Maximum Mean Discrepancy (MMD) on features from EEG-specific models (xDawn, EEGNet, REVE) alongside image-based ones. On the Reinforcement Learning and BCIAUT-P300 datasets, EEGGAN-SepH outperforms EEGGAN across nearly all fidelity and utility metrics and yields cleaner ERPs and spectra. A gaussian noise injection test further shows that the standard FD based on extracted features from time-frequency images is largely insensitive to mild noise, whereas EEG-specific metrics respond consistently, supporting their adoption for evaluating synthetic EEG.

Keywords: Generative Adversarial Networks, EEG, Synthetic Data, Evaluation Metrics

1 Introduction

Electroencephalography (EEG) is a non-invasive technique used to record brain activity by measuring the electrical exchanges between neurons

through dedicated electrodes positioned on the scalp [1]. EEG has many applications relevant not only to assisting disadvantaged individuals but also to supporting recovery, rehabilitation, and daily life more broadly, based on Motor Imagery (MI), sleep-stage classification, emotion recognition, recognition of Event-Related Potentials (ERPs), and the detection of neurological conditions such as Alzheimer’s disease, Parkinson’s disease, and autism spectrum disorder [2–4].

Deep Learning is becoming increasingly prevalent across systems able to support the life conditions of people, and it has also found application in EEG studies, achieving important results in pathology detection [5] as well as in assisting MI rehabilitation [6, 7].

Following these successes in classification, many architectures have been employed for other EEG-related tasks, with growing attention directed toward the generation field as well [8, 9].

Deep learning is used also to generate data, where generative models are effective in creating synthetic data with features similar to the ones of real EEG data [10]. Synthetic data offer a way to address several limitations present in the state of the art of EEG research, such as producing new datasets free from privacy issues, extending existing ones with new samples, and generating artificial patients sharing the same characteristics as real ones [11]. This leads to the use of data augmentation, enables larger-scale analyses

for clinicians, engineers, and AI experts, and supports the use of foundation models, which are known to be data-hungry.

Two families are currently prominent: Generative Adversarial Networks (GANs), where a generator improves through the feedback of a discriminator that judges the plausibility of the generated samples, and Diffusion Models, which learn to reverse a noising process until the signal takes the shape of an EEG trial [12]. Generative models are typically evaluated through qualitative inspection of EEG graphics and quantitative metrics that assess the fidelity of the generated data with respect to the real data such as Fréchet Inception Distance (FID) and CLIP MMD (CMMD) [13].

Many studies have already demonstrated the effectiveness of recreating synthetic EEGs that faithfully reproduce the principal components [14–16]; for this reason, deep learning-based generative models nowadays represent an innovative and highly relevant topic in the EEG field.

Although generative models offer significant benefits, the literature also highlights an important gap: synthetic EEG data often exhibit a high discrepancy, as measured by fidelity metrics, with respect to data recorded in a laboratory setting, which limits their practical use. Moreover, many studies rely on misleading metrics that fail to accurately capture this discrepancy. In this work, we address both gaps. On the generation side, we

extend EEGGAN [17] with the Separable Output Head (SepH), which converts the transformer output into the EEG trial through a spatial projection, a separable convolution and a learnable temporal smoothing filter. On the evaluation side, we compute the Fréchet Distance (FD) and the Maximum Mean Discrepancy (MMD) [18] on features from EEG-specific models, and we assess their reliability with a gaussian noise injection test.

The main contributions of this study are:

- a lightweight, EEG-oriented output head that models local temporal dynamics and suppresses high-frequency noise with only a few additional generator parameters.
- a pipeline in which FD and MMD on EEG-specific features respond consistently to gaussian noise, unlike the standard FID on time-frequency images.
- on two ERP datasets, EEGGAN-SepH yields cleaner ERPs and spectra, decreasing nearly all FD and MMD compared to the ones of EEGGAN; generating in the latent space of EEGGAN’s auxiliary autoencoder (AE) instead degrades fidelity.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on synthetic EEG data generation and evaluation. Section 3 describes the proposed SepH module, the training and the evaluation pipeline.

Section 4 details the datasets, preprocessing steps, experimental setup, and reports the experimental results. Section 5 presents the ablation study. Section 6 discusses the results, and Section 7 draws our conclusions.

2 Related work

2.1 Utility of Synthetic EEG Data

Since EEG datasets are typically small [19] and class-imbalanced [20], synthetic samples have proven effective for dataset extension [21]. Synthetic data yield classification gains on depressive disorder [22] and across EEG tasks [23], and even accuracy comparable to real data when training exclusively on synthetic samples for seizure prediction [24]. They can also support the pre-training of data-hungry foundation models [25] and reduce distribution shift across subjects, hardware and mental states [26]. Finally, they sidestep the privacy constraints that confine much of the literature to private, non-reproducible datasets [22], and have found use in federated learning and preprocessing analyses [26–28].

2.2 Generative Models for EEG

Early EEG data augmentation relied on training-free techniques such as mixup [29], SMOTE [30], gaussian noise injection [31], frequency shifting, time masking [32], channel shuffling [33], and Fourier transform surrogates [34], which merely

apply minor, often biologically implausible perturbations to real samples [35]. AEs and Variational AEs [36] only partially mitigated this, still producing near-copies of the training data. Most recent models therefore rely on GANs [37], commonly using the Wasserstein distance with gradient penalty (WGAN-GP) to prevent mode collapse [38, 39]. A class-conditioned WGAN-GP [15] improved the Gaussian Mixture Model (GMM)-based similarity metric [40] over GAN and WGAN-GP baselines, and BWGAN-GP [41] extended it by initializing discriminator and generator from a pre-trained autoencoder, so that generation starts from a learned latent embedding rather than random noise. EEGGAN adapts the Transformer Time Series GAN (TTS-GAN) [42] to EEG: a noise vector is linearly projected, positionally encoded, processed by transformer blocks, and mapped to the signal via a single 1×1 convolution, while the critic classifies embedded temporal patches through a prepended classification token; an optional transformer autoencoder can compress the signals to ease generation.

Diffusion models have also been explored. DESAM [43] mixes diffusion-generated and real motor imagery EEG to overcome the inconsistent results of fully synthetic augmentation, while U-Net-based diffusion models report FID and classification gains on motor imagery and

visual evoked potentials [16], and improved cross-subject generalization over GAN, DCGAN [44], and WGAN-GP baselines [45]. However, diffusion models require large training sets [46, 47], at odds with the scarce and imbalanced nature of EEG data [48], and their iterative sampling is poorly suited to real-time Brain Computer Interfaces (BCIs), whereas GANs generate samples in a single step.

More generally, these models often inherit architectures designed for generic time series or images, without an explicit focus on EEG, which limits their ability to capture its spatiotemporal relationships and results in poor fidelity. We address this gap by introducing the SepH for EEGGAN.

2.3 Evaluation Metrics for

Synthetic Signals

Since no single measure can characterize the plausibility of a generated biosignal, fidelity must be jointly assessed along temporal, spectral, statistical, and spatial dimensions. Prior work relies on time-domain errors and cross-correlation [49, 50], on spectral errors and comparisons of the Power Spectral Density [51], and on statistical divergences between the two distributions [52–54], whereas deep-learning-based metrics, such as the Fréchet Inception Distance (FID) [55], rely on features from networks pretrained on natural

images to assess realism and diversity [56, 57].

Spatial connectivity metrics, such as covariance and correlation distances, instead verify whether inter-electrode relationships are preserved [58, 59].

Nevertheless, most works adopt only a small subset of these measures, typically time-domain errors or features extracted by networks pre-trained on natural images, which overlook the physiological nature of brain activity [47]. They are often computed indiscriminately over noise, artifacts, and task-unrelated activity, and many fail to capture temporal shifts or detect mode collapse [60, 61]. We therefore adopt dedicated EEG-oriented fidelity metrics and feature extractors that better capture the physiologically plausible characteristics of the generated signals.

3 Methods

In this work, we present EEGGAN-SepH (Fig. 1), which uses an EEG-oriented output head for synthetic EEG generation built upon the EEGGAN [17]. The generator’s transformer encoder from random noise given as input produces a sequence of latent tokens, which are then converted into a synthetic EEG signal through the SepH step to preserve EEG features.

3.1 SepH module

While the transformer encoder in the EEGGAN effectively captures the temporal and spatial

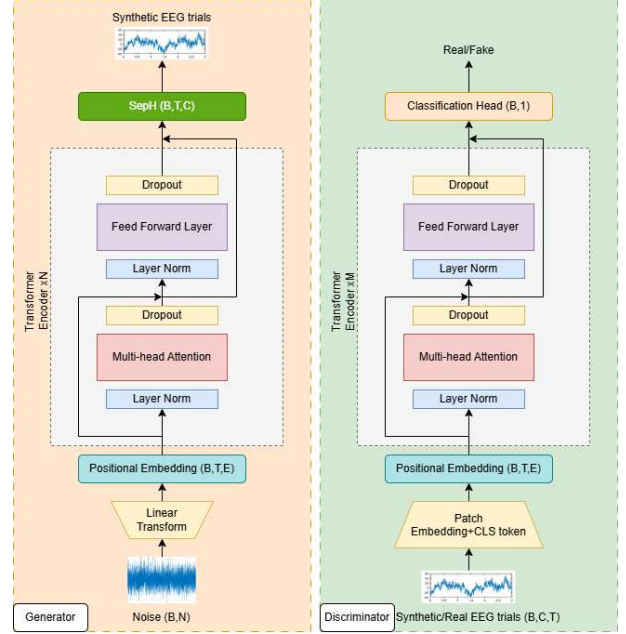


Fig. 1: Overview of the EEGGAN-SepH architecture. The generator (left) maps a noise vector to a sequence of latent tokens through a linear projection, positional embedding and N transformer encoder blocks; the proposed SepH then converts the tokens into the synthetic EEG trial. The critic (right) splits real or synthetic trials into temporal patches, prepends a classification token, processes them with M transformer encoder blocks and outputs a real/fake score through a classification head.

dynamics of the signal, it also introduces architectural bias and high-frequency noise into its output tokens. The simple single-step 2D convolution is not designed to remove such artifacts, resulting in synthetic signals affected by high-frequency noise and a distribution shift with respect to real EEG data.

To address this limitation, we replace the original output layer with an EEG-oriented module in the figure 2, inspired by EEGNet [62], which we

refer to as SepH. It departs from the original EEGNet block by introducing a spatial projection from the embedding space, replacing batch normalization with group normalization, using LeakyReLU instead of ELU, removing pooling to preserve the temporal resolution of the generated signal, and adding a learnable temporal smoothing filter.

Spatial projection.

The transformer output, of shape $(B, 1, T, E)$, where B is the batch, T is the sequence length and E the embedding dimension, is first projected through a 2D convolution with kernel size $(1, E)$, mapping the embedding dimension into $D \times C$ intermediate channels, where C is the number of EEG channels and D , following the convention introduced by EEGNet [62], denotes the number of spatial filters learned for each input channel. This yields a tensor of shape $(B, D \cdot C, T, 1)$, reintroducing an explicit spatial dimension. Group normalization, rather than batch normalization, is applied at this stage, followed by a LeakyReLU activation. Unlike batch normalization, which computes statistics across the batch dimension and can thus introduce unwanted dependencies between samples during generation, group normalization computes statistics independently for each sample, avoiding this cross-sample contamination.

Separable convolution.

Inspired by EEGNet’s depthwise-separable block, the projected tensor is processed by a depthwise convolution (kernel size 3, groups = $D \cdot C$), which operates along the temporal dimension, learning local temporal dynamics independently for each of the $D \cdot C$ channels. This is followed by a pointwise convolution (kernel size 1), which instead operates across the $D \cdot C$ dimension, mixing information between channels to capture spatial dependencies and produce the final C -channel output.

Temporal smoothing.

To suppress residual high-frequency noise, a depthwise, learnable temporal smoothing filter is applied independently to each channel, with kernel size adaptively set relative to the sequence length and initialized from a gaussian kernel. To avoid over-smoothing, which could remove physiologically relevant high-frequency activity, the smoothed signal x_{smooth} is combined with the pre-smoothing signal x through a learnable residual mix:

$$x' = \alpha x_{\text{smooth}} + (1 - \alpha) x, \quad (1)$$

where $\alpha \in [0, 1]$ is learned during training, allowing the model to adaptively balance noise suppression against signal preservation.

3.2 EEGGAN-SepH Training

EEGGAN-SepH is trained using the Wasserstein loss with a gradient penalty regularization term,

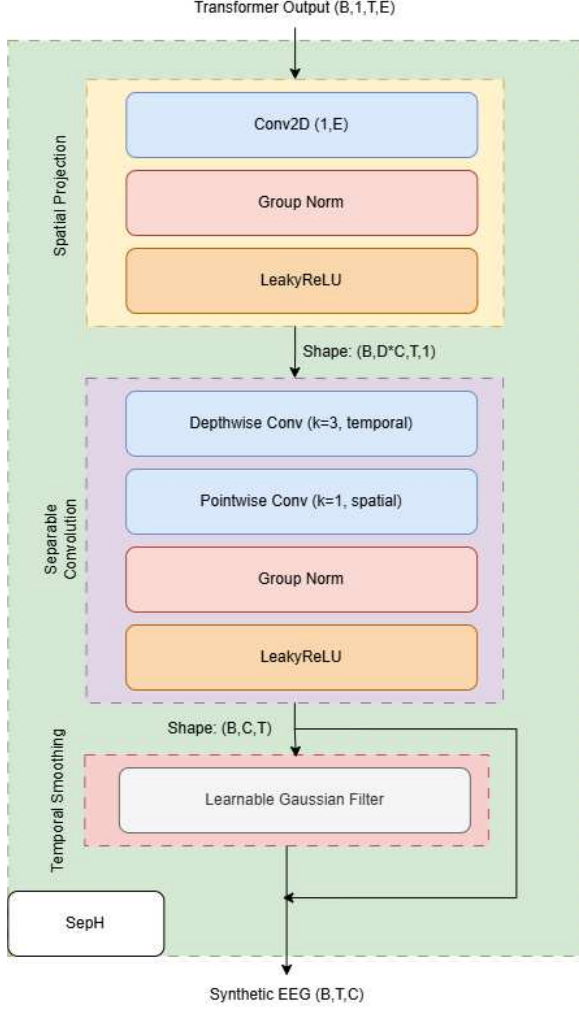


Fig. 2: Schematic of the proposed SepH module. The transformer output is first processed by a spatial projection, then by a separable convolution block, and finally by a learnable temporal smoothing filter, whose output is combined with the pre-smoothing signal through a residual mix.

which prevents mode collapse and stabilizes training [63]. The critic loss is defined as:

$$L_D = \mathbb{E}_{\hat{x} \sim \mathbb{P}_g} [D(\hat{x})] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)] + \lambda_{gp} \mathbb{E}_{\hat{x} \sim \mathbb{P}_g} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \quad (2)$$

while the generator loss is defined as:

$$L_G = -\mathbb{E}_{\hat{x} \sim \mathbb{P}_g} [D(\hat{x})], \quad (3)$$

where $\mathbb{P}_r$ and $\mathbb{P}_g$ denote the real and generated data distributions, respectively, $\hat{x}$ is a sample obtained via linear interpolation between real and generated data, and λ_{gp} is the gradient penalty coefficient. The model is class-conditioned by appending the labels as additional input channels; on Dataset 2, it is conditioned also for subject and session.

3.3 Evaluation Metrics

We evaluate the synthetic samples through both a qualitative and a quantitative assessment that complements commonly used image-based metrics with features from EEG-specific models. The qualitative evaluation inspects the trial-averaged Event-Related Potential (ERP), frequency spectrum, and channel-wise correlation matrix, while the quantitative one compares the feature distributions of real (P) and synthetic (Q) samples through the Fréchet Distance (FD, referred to as FID when computed on Inception-v3 features) and the Maximum Mean Discrepancy (MMD), both computed on the embeddings that a pretrained or trained feature extractor produces for each trial.

FD is the Wasserstein-2 distance between the multivariate Gaussians $\mathcal{N}(\mu_P, \Sigma_P)$ and

$\mathcal{N}(\mu_Q, \Sigma_Q)$ fitted to the real and synthetic embeddings:

$$\text{FD}(P, Q) = \|\mu_P - \mu_Q\|_2^2 + \text{Tr} \left(\Sigma_P + \Sigma_Q - 2(\Sigma_P \Sigma_Q)^{1/2} \right), \quad (4)$$

where the covariance matrices are estimated by Maximum Likelihood Estimation (MLE) for Inception-v3 features and by Oracle Approximating Shrinkage (OAS) for all other extractors, to ensure numerical stability. MMD is a kernel-based, non-parametric two-sample statistic measuring the distance between the mean embeddings of P and Q in a reproducing kernel Hilbert space:

$$\text{MMD}^2(P, Q) = \mathbb{E}[k(\mathbf{x}, \mathbf{x}')] + \mathbb{E}[k(\mathbf{y}, \mathbf{y}')] - 2\mathbb{E}[k(\mathbf{x}, \mathbf{y})], \quad (5)$$

with $\mathbf{x}, \mathbf{x}' \sim P$, $\mathbf{y}, \mathbf{y}' \sim Q$, and $k(\cdot, \cdot)$ a Radial Basis Function (RBF) kernel with bandwidth $\sigma = 10$; reported values are scaled by a factor of 10^3 . For both metrics, lower values indicate closer agreement between real and synthetic distributions.

Since FD and MMD compare embeddings rather than raw signals, their outcome depends on the network used to extract them. We compute the metrics over the features extracted by the following extractors: xDAWN [64] provides spatial filters that enhance the evoked response, EEGNet [62]

provides task-discriminative features learned on the same data, and REVE [65] provides general-purpose representations from a foundation model pretrained on large-scale EEG, while Inception-v3 [66] and CLIP [67] serve as the image-based references commonly adopted in the literature. The two image-based models receive a grayscale Short-Time Fourier Transform (STFT) time-frequency image of each trial, resized and normalized according to the input requirements of each network. The EEG-based models operate directly on the trials: REVE is used as a frozen pretrained encoder, xDAWN is fitted on the training set of each dataset, and EEGNet is trained with supervision on the real training set over three random seeds. For REVE and EEGNet, features are extracted from the penultimate layer.

For Dataset 1, we generate 5000 synthetic samples per class and compute both metrics on all real and synthetic samples. For BCIAUT-P300 Dataset, 5000 trials are randomly drawn from each of the training, test and synthetic sets, pooling subjects and sessions. Although the generator is conditioned on class, subject and session, the metrics could not be computed at the session level: each session contains only 200 target trials, which is insufficient to reliably estimate the covariance matrices required by the FD and makes the metric unstable. In both cases, the metrics are computed against both the training without mixup and the test set.

To assess the utility of the synthetic data, we evaluate them in a transfer-learning setting on Dataset 2, which is based on an oddball paradigm with a 1:7 target-to-non-target ratio. For each subject and session, EEGNet was first pre-trained on a set derived from all subjects and sessions of Dataset 2 except the session used for fine-tuning, and then fine-tuned on the calibration data of that session. Four pre-training strategies were compared: no pre-training, pre-training on real data, and pre-training on data generated by EEGGAN or EEGGAN-SepH. To evaluate the impact of the amount of fine-tuning data, the calibration set was reduced proportionally to 12.5%, 25%, 50%, and 100% of its size, preserving the class ratio. Performance was measured on held-out data from the same session using balanced accuracy and the area under the ROC curve (AUC). In both pre-training and fine-tuning, EEGNet was trained with a class-weighted loss to account for class imbalance, and each configuration was repeated with three random seeds.

To assess the sensitivity of each metric, we additionally perform a noise injection test on the training set of Dataset 2: gaussian noise with standard deviation increasing from 0 to 1 in steps of 0.1 is added to the z-scored real trials, and FD and MMD are computed between the clean and the noisy sets for each class and feature extractor. A reliable metric is expected to increase monotonically with the noise level.

4 Experiments and Results

4.1 Dataset

To evaluate the proposed framework, we conduct experiments on two EEG datasets, both representing event-related potentials (ERPs) elicited by visual stimuli.

(1) Reinforcement Learning Dataset

This dataset captures neural responses to binary win/loss visual feedback during a two-armed bandit task [68]. EEG data were recorded at 500 Hz using a 32- or 64-electrode ActiCAP system (Brain Products GmbH, Munich, Germany). The training set comprises 15 participants and 1129 samples (516 Win, 613 Lose); the test set comprises 200 additional participants and 15516 samples (7536 Win, 7980 Lose).

(2) BCIAUT-P300 Dataset

BCIAUT-P300 [69] is a multi-session P300-based BCI dataset from 15 individuals with autism spectrum disorder undergoing joint-attention training, comprising 105 sessions (7 per participant). Participants visually identified a target object among flashed distractors, eliciting a target-related P300 response when the attended object matched the requested one. Each session includes a calibration phase (train) and an online phase (test): calibration consists of 20 blocks of 10 runs, yielding 1600 epochs per session (200 target, 1400 non-target) and was used to determine, per session, the minimum number of runs per block

(K) achieving above 80% classifier accuracy; the online phase then used 50 blocks with K (ranging 3–10) runs each. EEG data were recorded at 250 Hz from eight channels (C3, Cz, C4, CPz, P3, Pz, P4, POz), referenced at the right ear with ground at AFz.

4.2 Data Preprocessing

For RL Dataset, data were preprocessed following [70]. Only the channel FCz is selected. Signals were downsampled to 100 time points (effective sampling rate of 83.33 Hz) and epoched from -200 ms to 1000 ms relative to feedback onset. Min-max normalization to the $[0, 1]$ range was applied prior to training, following the same prior work.

For BCIAUT-P300 Dataset, data were resampled from 250 Hz to 128 Hz and epoched from -200 ms to 800 ms relative to stimulus onset. This window was selected based on inspection of the grand-average ERP across all subjects and sessions, which showed that neural activity returned to pre-stimulus baseline levels by 800 ms. A gaussian filter (kernel size 13, $\sigma = 2.5$) was applied to both training and test sets to suppress high-frequency noise while preserving the low-frequency content where most of the P300 power lies, which is what we want to recreate in the synthetic data. To address the strong class imbalance between target and non-target trials, a mixup [29] operation was applied to the training set, combining trials belonging to the target class,

same session, and subject to generate new samples. Mixup is used only to train the generative models and not on metric evaluation. Trial-wise standardization (zero mean, unit standard deviation) was applied to reduce variability across sessions and subjects.

4.3 Device and Parameters

Experiments were implemented in PyTorch and run on an Intel Core i9-14900K CPU with an NVIDIA GeForce RTX 4090 GPU (24 GB).

GAN hyperparameters mostly follow [17]. On both datasets, generator and critic use four transformer layers with five attention heads, Adam ($\beta_1 = 0$, $\beta_2 = 0.9$, learning rate 10^{-4} for both), a latent dimension of 128, $\lambda_{gp} = 10$, and critic dropout and forward dropout of 0.5. On Dataset 1, the GAN is trained for 4,000 epochs with batch size 128, five critic iterations per generator update, generator forward and attention dropout of 0.5, generator and critic embedding dimensions of 10 and 50, sequence length 100, and class conditioning; the patch size is 10 in the AE latent space and 20 on raw signals (five patches in both cases). On Dataset 2, it is trained for 2,000 epochs with batch size 2,048, two critic iterations per generator update, generator forward and attention dropout of 0.2, embedding dimensions of 60 and 120, sequence length 128, patch size 16, and conditioning on class, subject, and session. The critic

iterations were reduced because the critic otherwise overpowered the generator, and the dropout was lowered to stabilize training; the larger batch size speeds up training, with embedding dimensions scaled accordingly to keep the two networks balanced.

The auxiliary AE, used only on Dataset 1, compresses the signals along time to a length of 50 using a two-layer transformer (hidden dimension 256, eight attention heads, dropout 0.1) with a sigmoid decoder output, trained for 2,000 epochs with batch size 128, learning rate 10^{-4} and training ratio 0.8.

On both datasets, the SepH uses $D = 2$, a $(1, E)$ spatial projection kernel, with E the embedding dimension, a depthwise convolution with kernel size 3 and padding 1, a pointwise convolution with kernel size 1, group normalization after both the spatial projection and the channel mixing, and LeakyReLU (negative slope 0.2). The temporal smoothing kernel has size 7 on Dataset 1 and 9 on Dataset 2, and the mixing coefficient α , initialized to 0.5, converges to 0.74 and 0.89, respectively. On Dataset 1, the generator has 136,331 parameters in EEGGAN and 136,362 in EEGGAN-SepH (+31); on Dataset 2, 1,197,848 and 1,198,617 (+769). The critic is unchanged (125,151 and 720,241 parameters on the two datasets), and the AE has 2,452,118 parameters.

4.4 Results and Analysis

To qualitatively evaluate the proposed model, we compare the z-scored grand-average ERPs, averaged across subjects, of real and synthetic data. On Dataset 1 (Fig. 3), we compare four configurations: EEGGAN, EEGGAN operating in the latent space of the AE (EEGGAN-AE), EEGGAN-SepH, and its AE-based variant (EEGGAN-AE-SepH). The ERPs generated by EEGGAN-SepH more closely resemble the Train and Test ERPs than those generated by EEGGAN-AE. On Dataset 2 (Fig. 4), the ERPs generated by EEGGAN-SepH closely match the real ones and show fewer spikes, a typical symptom of the high-frequency noise introduced by generative models. Consistently, the frequency spectrum computed via the Fast Fourier Transform (FFT) reveals less high-frequency noise in the data generated by the proposed model (Fig. 5). Finally, the grand-average inter-channel correlation is preserved by both architectures (Fig. 6).

In the quantitative results for Dataset 1, the Train vs Test comparison reports among the lowest scores across most fidelity metrics for both classes (Table 1), and serves as the reference discrepancy against which the generative models are assessed.

Comparing EEGGAN against EEGGAN-SepH shows a consistent improvement across all metrics for both classes and both reference sets. Against

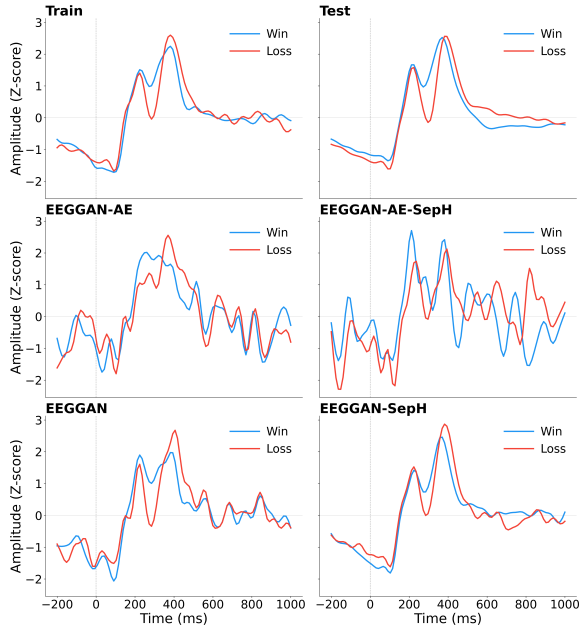


Fig. 3: Grand-average ERP waveforms for the Win and Lose conditions on RL Dataset. A comparison between real data (Train, Test) against synthetic data generated by four models: EEGGAN-AE with the original output head, EEGGAN-AE with the proposed SepH module, EEGGAN, and EEGGAN-SepH.

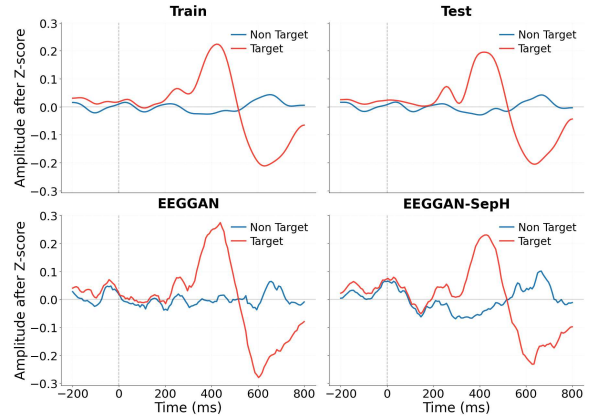


Fig. 4: Grand-average ERP waveforms for the Target and Non-Target conditions on BCIAUT-P300 Dataset. Z-scored ERPs averaged across subjects and sessions of the BCIAUT-P300 dataset for real data (Train and Test, top row) and for synthetic data generated by EEGGAN with the original output head and by EEGGAN-SepH (bottom row). Time 0 ms corresponds to stimulus onset. The ERPs generated by EEGGAN-SepH closely match the real ones and show fewer high-frequency spikes than those generated by EEGGAN.

Test, FD xDAWN drops by roughly two thirds (from 0.4999 to 0.1677 for Win) and FD EEG-Net by more than half (from 0.1297 to 0.0569), with comparable reductions for the Lose class and against Train. The improvement is confirmed across MMD xDAWN, CMMD, MMD REVE and MMD EEGNet, and, notably, also by FD Inception: in this direct comparison all feature extractors agree on the benefit of the proposed output head. The smallest gain, for MMD EEGNet on the Lose class against Test, lies within the seed-to-seed standard deviation.

This picture changes once the auxiliary autoencoder is included: EEGGAN-AE-SepH *underperforms* EEGGAN-AE across nearly all metrics and both reference sets, with FD xDAWN roughly doubling for both classes. For the Win class the reversal holds across all eight metrics; for the Lose class, CMMD is the only exception. Comparing EEGGAN-SepH against EEGGAN-AE-SepH directly, i.e., isolating the effect of the autoencoder from that of the output head, FD xDAWN, FD REVE, FD EEGNet and FD Inception all consistently favor EEGGAN-SepH (e.g., FD EEGNet of 0.0569 vs. 0.3002 for Win against Test). Combining the auxiliary autoencoder with

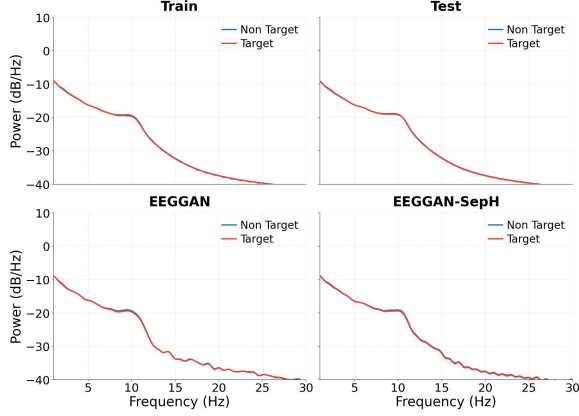


Fig. 5: Grand-average of the frequency spectrum for the Target and Non-Target conditions on BCIAUT-P300 Dataset. Power spectral density, computed via the Fast Fourier Transform, for real data (Train and Test, top row) and for synthetic data generated by EEGGAN with the original output head and by EEGGAN-SepH (bottom row) on the BCIAUT-P300 dataset. The spectrum of EEGGAN-SepH shows less high-frequency noise than that of EEGGAN.

the proposed module therefore has a negative effect on fidelity. FD Inception for EEGGAN-SepH against Test nevertheless remains well above the Train vs Test reference, indicating that, under Inception features, the synthetic data are still less similar to the real test set than the two real splits are to each other.

Adding a gaussian filter to the output of EEGGAN and EEGGAN-AE improves the Inception- and CLIP-based scores in every setting, to the point that EEGGAN-AE + Gauss filt. reaches the lowest FD Inception and CMMD among all generative models. The xDAWN-based scores also decrease, whereas the learned EEG-domain extractors do not support this improvement:

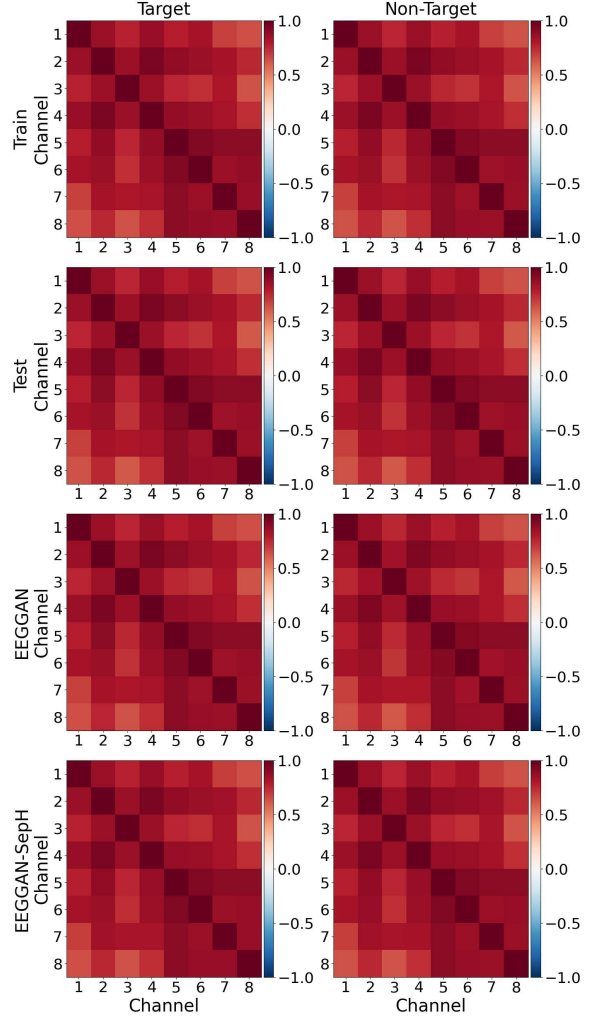


Fig. 6: Grand-average of the correlation between channels for the Target and Non-Target conditions on BCIAUT-P300 Dataset. Inter-channel correlation matrices averaged across trials of the BCIAUT-P300 dataset for real data (Train and Test) and for synthetic data generated by EEGGAN with the original output head and by EEGGAN-SepH (one row each). Both architectures preserve the inter-channel correlation structure of the real data.

MMD EEGNet worsens in every setting, and both gaussian-filtered models remain far from EEGGAN-SepH on EEG-specific features (FD EEGNet of 0.1305 and 0.1414 against 0.0569 for

Win against Test). Notably, in the ablation study of Section 5, EEGGAN-SepH without the learnable filter obtains lower FD Inception and CMMD against Test than EEGGAN-AE + Gauss filt. for both classes.

Since the FD relies on covariance matrix estimation, whose accuracy depends on the number of samples in each set, and since Test contains more samples than Train, FD scores computed against Test tend to be lower than those computed against Train for the same generative model; the two references should therefore be compared only within themselves, as indicated in Table 1. This effect does not concern MMD, which does not require covariance estimation. For both classes, MMD xDAWN between Train and EEGGAN-SepH is equal to or lower than the Train vs Test reference, which we attribute to the limitations of this metric discussed below.

On Dataset 2, the improvement introduced by the SepH is consistent across nearly all metrics for both classes (Table 2). Against Train, FD xDAWN decreases by roughly 43% (from 6.9431 to 3.9937 for Non-Target) and FD EEGNet by about one third, while MMD EEGNet and MMD REVE show the largest relative improvements, decreasing by roughly twofold and threefold respectively; the same pattern is confirmed against Test. FD Inception also improves in both reference sets, in agreement with the EEG-oriented metrics. The one exception is MMD xDAWN, which increases with

EEGGAN-SepH for both classes against Train and for the Target class against Test, remaining essentially unchanged for Non-Target. We note that MMD EEGNet for EEGGAN carries a large seed-to-seed standard deviation, so the corresponding improvement should be read with caution.

Adding a gaussian filter to the output of EEGGAN leads to a markedly different pattern on Dataset 2. The xDAWN- and REVE-based scores decrease in nearly all settings, and EEGGAN + Gauss filt. obtains the lowest MMD xDAWN in three of the four comparisons. In contrast, FD Inception and CMMD worsen in every setting (e.g., FD Inception from 1.6156 to 3.3780 for Test, Non-Target), i.e., in the opposite direction to Dataset 1, where the same operation yielded the best scores on these two metrics. MMD EEGNet worsens in every setting, as on Dataset 1, and FD EEGNet does not improve either, its only decrease lying within the seed-to-seed standard deviation.

Overall, EEGGAN-SepH outperforms EEGGAN + Gauss filt. on all metrics except MMD xDAWN.

That one and the same operation, low-pass filtering the generated signals, substantially improves Inception and CLIP on one dataset and degrades them on the other, while the EEG-specific extractors respond in the same direction on both, is direct evidence that image-based features are not reliable for assessing the fidelity of EEG signals.

Table 3 shows that pre-training on data generated by EEGGAN-SepH improves the classification results obtained with data generated by EEGGAN, with balanced accuracy increasing from 64.75% to 68.66% with 25 targets / 175 non-targets and from 71.71% to 73.20% with 200 targets / 1400 non-targets, and AUC from 70.12% to 75.07% and from 78.43% to 80.20%, respectively. In particular, pre-training on EEGGAN-SepH synthetic data is most helpful when fewer fine-tuning data are available: with 25 targets / 175 non-targets, balanced accuracy rises from 62.58% without pre-training to 68.66% (AUC from 67.38% to 75.07%), and with 50 targets / 350 non-targets from 67.46% to 70.14%. This benefit decreases as the calibration set grows, becoming marginal with 100 targets / 700 non-targets (71.71% vs. 71.19%) and reversing with 200 targets / 1400 non-targets (73.20% vs. 74.07%). Pre-training on EEGGAN data, instead, is beneficial only with 25 targets / 175 non-targets and falls below training from scratch for all larger budgets (e.g., 71.71% vs. 74.07% with 200 targets / 1400 non-targets).

Nevertheless, pre-training on real data still yields the best results for every budget, with balanced accuracy of 71.37%, 72.49%, 73.84%, and 75.19%, compared to 68.66%, 70.14%, 71.71%, and 73.20% for EEGGAN-SepH.

Fig. 7 shows the training set of Dataset 2 evaluated with FD and MMD on xDAWN, REVE, EEGNet and Inception/CLIP features, as a function of the gaussian noise added to the training samples, with standard deviation increasing in steps of 0.1. Overall, all metrics respond consistently to the amount of injected noise. MMD with xDAWN is the most problematic metric in this case, since it slightly decreases at low noise levels, which may bias the evaluation on Dataset 2; the Fréchet distance on the same features does not suffer from this issue. The results show that FID on images given by a STFT is almost insensitive to low noise, which suggests it is not the most suitable metric for evaluating synthetic EEG data. The results obtained with REVE and EEGNet are clearly more reliable.

Table 1: FD and MMD scores on RL Dataset for Win and Lose classes.

		FD ↓				MMD ↓			
Class	Comparison	xDAWN	Inception	REVE	EEGNet	xDAWN	CLIP	REVE	EEGNet
Reference									
Win	Train vs Test	0.0626	8.6121	0.0058	0.0248 ± 0.0043	0.0010	0.0017	0.0032	0.0153 ± 0.0003
Lose	Train vs Test	0.0522	7.6002	0.0052	0.0212 ± 0.0038	0.0011	0.0010	0.0039	0.0220 ± 0.0037
Train-based comparisons									
Win	EEGGAN	0.5198	30.3015	0.0273	0.1433 ± 0.0227	0.0035	0.1647	0.0380	0.0890 ± 0.0156
	EEGGAN + Gauss filt.	0.4178	28.9777	0.0283	0.1528 ± 0.0284	0.0014	0.1148	0.0403	0.1206 ± 0.0324
	EEGGAN-AE	0.5423	30.7969	0.0328	0.1763 ± 0.0223	0.0079	0.0899	0.0762	0.1953 ± 0.0289
	EEGGAN-AE + Gauss filt.	0.4888	22.5803	0.0298	0.1564 ± 0.0205	0.0049	0.0578	0.0538	0.1988 ± 0.0145
	EEGGAN-SepH	0.2290	27.2424	0.0152	0.0773 ± 0.0114	0.0010	0.0978	0.0232	0.0589 ± 0.0221
	EEGGAN-AE-SepH	0.9592	36.6599	0.0443	0.2998 ± 0.0285	0.0279	0.1124	0.1450	0.7265 ± 0.0650
Lose	EEGGAN	0.5357	31.1540	0.0286	0.1462 ± 0.0242	0.0050	0.1683	0.0455	0.1090 ± 0.0275
	EEGGAN + Gauss filt.	0.4088	28.6613	0.0276	0.1514 ± 0.0298	0.0017	0.1132	0.0497	0.1210 ± 0.0421
	EEGGAN-AE	0.5505	28.2599	0.0316	0.1720 ± 0.0209	0.0088	0.0850	0.0666	0.1972 ± 0.0084
	EEGGAN-AE + Gauss filt.	0.4849	20.3681	0.0270	0.1507 ± 0.0202	0.0055	0.0536	0.0485	0.2191 ± 0.0166
	EEGGAN-SepH	0.2150	26.6978	0.0134	0.0787 ± 0.0132	0.0008	0.1004	0.0201	0.0870 ± 0.0298
	EEGGAN-AE-SepH	1.0238	31.0913	0.0555	0.3090 ± 0.0225	0.0257	0.0765	0.1901	0.7548 ± 0.0690
Test-based comparisons									
Win	EEGGAN	0.4999	25.5776	0.0255	0.1297 ± 0.0211	0.0044	0.1625	0.0449	0.0992 ± 0.0141
	EEGGAN + Gauss filt.	0.3546	20.8308	0.0229	0.1305 ± 0.0254	0.0019	0.1061	0.0365	0.1287 ± 0.0308
	EEGGAN-AE	0.5712	23.6455	0.0318	0.1722 ± 0.0208	0.0086	0.0955	0.0863	0.1993 ± 0.0294
	EEGGAN-AE + Gauss filt.	0.4647	13.2558	0.0260	0.1414 ± 0.0166	0.0054	0.0477	0.0575	0.2056 ± 0.0169
	EEGGAN-SepH	0.1677	18.8917	0.0102	0.0569 ± 0.0084	0.0018	0.0923	0.0207	0.0666 ± 0.0211
	EEGGAN-AE-SepH	0.9913	29.7378	0.0445	0.3002 ± 0.0272	0.0297	0.1217	0.1606	0.7730 ± 0.0705
Lose	EEGGAN	0.5236	28.6510	0.0256	0.1351 ± 0.0217	0.0058	0.1724	0.0445	0.1197 ± 0.0245
	EEGGAN + Gauss filt.	0.3668	19.2364	0.0251	0.1324 ± 0.0256	0.0021	0.1007	0.0503	0.1288 ± 0.0424
	EEGGAN-AE	0.5810	25.1147	0.0300	0.1737 ± 0.0202	0.0091	0.0951	0.0732	0.2015 ± 0.0077
	EEGGAN-AE + Gauss filt.	0.4804	12.2286	0.0252	0.1429 ± 0.0167	0.0060	0.0464	0.0563	0.2270 ± 0.0178
	EEGGAN-SepH	0.1892	17.5455	0.0103	0.0642 ± 0.0094	0.0015	0.0896	0.0211	0.1077 ± 0.0307
	EEGGAN-AE-SepH	1.0730	28.4765	0.0542	0.3117 ± 0.0184	0.0272	0.0874	0.2053	0.8001 ± 0.0445

FD is computed on xDAWN, Inception-v3, REVE and EEGNet features; MMD on xDAWN, CLIP, REVE and EEGNet features.

Train vs Test reports the reference discrepancy between the two real-data splits. Bold values indicate the lowest score for each metric, separately within each class and within the Train-based and Test-based comparisons. *EEGGAN*: original output head; *EEGGAN-SepH*: proposed SepH module; *EEGGAN-AE*: autoencoder-based variant; + *Gauss filt.*: addition of gaussian filtering.

EEGNet-based scores are mean $\pm$ std over 3 random seeds; other scores refer to a single run.

Table 2: FD and MMD scores on BCIAUT-P300 Dataset for Target and Non-Target classes.

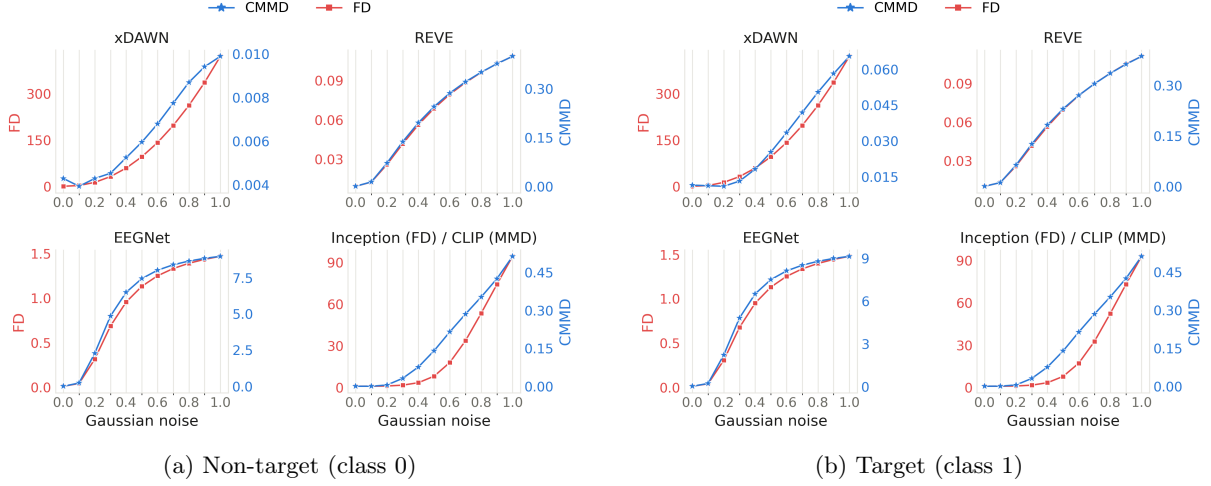


Fig. 7: FD and MMD as a function of gaussian noise on BCIAUT-P300 Dataset. Gaussian noise with standard deviation increasing from 0 to 1 in steps of 0.1 is added to the z-scored real trials of the BCIAUT-P300 training set, and FD and MMD are computed between the clean and the noisy sets for the non-target (a) and target (b) classes. Each panel refers to a different feature extractor: xDAWN, REVE, EEGNet and FID / CMMD. FD is shown on the left axis (red) and MMD on the right axis (blue, labelled CMMD in the panels). A reliable metric is expected to increase monotonically with the noise level.

		FD ↓				MMD ↓			
Class	Comparison	xDAWN	Inception	REVE	EEGNet	xDAWN	CLIP	REVE	EEGNet
Reference									
Non-Target	Train vs Test	0.7854	0.9719	0.0094	0.0118 ± 0.0003	0.0043	0.0008	0.0008	0.0123 ± 0.0012
Target	Train vs Test	0.8821	0.9564	0.0103	0.0156 ± 0.0006	0.0116	0.0010	0.0012	0.0360 ± 0.0037
Train-based comparisons									
Non-Target	EEGGAN	6.9431	1.4609	0.0165	0.0714 ± 0.0104	0.0161	0.0024	0.0104	0.1189 ± 0.0854
	EEGGAN + Gauss filt.	4.8433	2.5945	0.0158	0.0734 ± 0.0020	0.0145	0.0082	0.0057	0.1538 ± 0.0068
	EEGGAN-SepH	3.9937	1.2364	0.0140	0.0489 ± 0.0013	0.0209	0.0011	0.0036	0.0500 ± 0.0041
Target	EEGGAN	6.3979	1.2365	0.0167	0.0611 ± 0.0134	0.0179	0.0021	0.0095	0.1285 ± 0.1061
	EEGGAN + Gauss filt.	4.1861	2.5190	0.0159	0.0594 ± 0.0012	0.0141	0.0069	0.0048	0.1385 ± 0.0043
	EEGGAN-SepH	3.5209	1.0875	0.0144	0.0411 ± 0.0015	0.0190	0.0004	0.0030	0.0404 ± 0.0060
Test-based comparisons									
Non-Target	EEGGAN	7.4946	1.6156	0.0170	0.0682 ± 0.0078	0.0170	0.0039	0.0114	0.0854 ± 0.0663
	EEGGAN + Gauss filt.	5.6545	3.3780	0.0165	0.0826 ± 0.0020	0.0174	0.0120	0.0068	0.2223 ± 0.0045
	EEGGAN-SepH	4.3598	1.4075	0.0144	0.0504 ± 0.0019	0.0170	0.0018	0.0033	0.0460 ± 0.0080
Target	EEGGAN	7.0443	1.3941	0.0171	0.0619 ± 0.0106	0.0271	0.0030	0.0104	0.1101 ± 0.0821
	EEGGAN + Gauss filt.	5.1045	3.3341	0.0169	0.0756 ± 0.0009	0.0243	0.0104	0.0072	0.2505 ± 0.0110
	EEGGAN-SepH	3.9391	1.3189	0.0147	0.0461 ± 0.0013	0.0287	0.0012	0.0041	0.0625 ± 0.0046

FD is computed on xDAWN, Inception-v3, REVE and EEGNet features; MMD on xDAWN, CLIP, REVE and EEGNet features.

Train vs Test reports the reference discrepancy between the two real-data splits. Bold values indicate the lowest score for each

metric, separately within each class and within the Train-based and Test-based comparisons. *EEGGAN*: original output head; *EEGGAN-SepH*: proposed SepH module; + *Gauss filt.*: addition of gaussian filtering. EEGNet-based scores are mean $\pm$ std over 3 random seeds; other scores refer to a single run.

Table 3: Cross-session classification with pretraining on different configurations for BCIAUT-P300 Dataset.

Pre-training strategy	Balanced accuracy [%] $\uparrow$				AUC [%] $\uparrow$			
	25 / 175	50 / 350	100 / 700	200 / 1400	25 / 175	50 / 350	100 / 700	200 / 1400
From scratch	62.58 $\pm$ 0.36	67.46 $\pm$ 0.42	71.19 $\pm$ 0.48	74.07 $\pm$ 0.16	67.38 $\pm$ 0.49	73.46 $\pm$ 0.44	77.98 $\pm$ 0.44	81.22 $\pm$ 0.18
EEGGAN	64.75 $\pm$ 0.48	67.07 $\pm$ 0.49	69.57 $\pm$ 0.30	71.71 $\pm$ 0.30	70.12 $\pm$ 0.56	72.94 $\pm$ 0.59	75.99 $\pm$ 0.37	78.43 $\pm$ 0.34
EEGGAN-SepH	68.66 $\pm$ 0.20	70.14 $\pm$ 0.22	71.71 $\pm$ 0.04	73.20 $\pm$ 0.05	75.07 $\pm$ 0.28	76.71 $\pm$ 0.30	78.62 $\pm$ 0.11	80.20 $\pm$ 0.16
Real EEG	71.37 $\pm$ 0.28	72.49 $\pm$ 0.09	73.84 $\pm$ 0.14	75.19 $\pm$ 0.11	78.41 $\pm$ 0.07	79.62 $\pm$ 0.11	81.10 $\pm$ 0.19	82.68 $\pm$ 0.12

For each session, EEGNet is pre-trained on a pool that excludes the calibration session and then fine-tuned on an increasing amount of real calibration data. Column headers report the number of calibration targets / non-targets, obtained by scaling the full session (200 targets / 1400 non-targets) by a factor of 0.125, 0.25, 0.5 and 1.0, so that the target-to-non-target ratio is preserved. Performance is evaluated on held-out data from the same session. *From scratch*: no pre-training; *EEGGAN*: pre-training on data generated with the original output head; *EEGGAN-SepH*: pre-training on data generated with the proposed SepH module; *Real EEG*: pre-training on real data only, acting as the upper bound. All values are percentages, reported as the mean over 3 random seeds $\pm$ standard deviation, each seed being averaged over 15 subjects $\times$ 7 sessions. Bold values indicate the best score for each column.

5 Ablation Study

We conducted an ablation study on Dataset 1 (Table 4) to assess how the individual components of the SepH module affect performance, modifying one component at a time.

To evaluate the contribution of Group Normalization, we replaced it with Batch Normalization. This change consistently worsens all metrics for both classes, with MMD EEGNet increasing from 0.0666 to 1.6301 for Win against Test, indicating that Group Normalization is essential for fidelity.

To evaluate the role of the learnable temporal smoothing filter, we removed it from the module. This worsens the xDAWN and EEGNet-based metrics for both classes (e.g., FD EEGNet from

0.0569 to 0.1564 for Win against Test), confirming that the filter improves fidelity in EEG-specific feature spaces. In contrast, removing it improves the metrics based on models pretrained on natural images (FD Inception: from 18.8917 to 9.7178 and from 17.5455 to 11.8547; CMMD: from 0.0923 to 0.0378 and from 0.0896 to 0.0380, for Win and Lose against Test), as well as MMD REVE.

We conducted an ablation study on Dataset 2 (Table 5) to assess the contribution of the point-wise convolution within the separable block. The results are remarkably balanced: the two configurations split the wins on FD xDAWN, FD Inception, and FD EEGNet, and the differences on the latter (e.g., 0.0504 vs. 0.0478 for Non-Target

against Test) are comparable to the seed-to-seed CLIP). The learnable filter, instead, improves standard deviation. Nevertheless, a slight class-dependent trend emerges: for the Target class, which contains the P300 response, EEGAN-SepH obtains the lower score in five of the eight metrics with respect to both Train and Test, whereas for the Non-Target class the same holds for EEGAN-SepH-CW.

Overall, the ablation study shows that replacing the batch normalization used in EEGNet with group normalization improves fidelity across all metrics, including those computed on generic feature extractors (Inception-v3,

most EEG-specific metrics but slightly worsens those based on generic feature extractors, revealing a discrepancy between EEG-specific and generic feature spaces that does not emerge for group normalization. Cross-electrode mixing yields fidelity comparable to a channel-wise convolution, indicating that the benefit of SepH does not stem primarily from inter-electrode exchange, although it may help model the spatial structure of the P300.

Table 4: Ablation of the SepH components on RL Dataset.

		FD ↓				MMD ↓			
Class	Model	xDAWN	Inception	REVE	EEGNet	xDAWN	CLIP	REVE	EEGNet
Train-based comparisons									
Win	EEGGAN-SepH	0.2290	27.2424	0.0152	0.0773 ± 0.0114	0.0010	0.0978	0.0232	0.0589 ± 0.0221
	EEGGAN-SepH w/o Group Norm	1.3385	44.0625	0.0187	0.2872 ± 0.0659	0.0080	0.1224	0.0300	1.6389 ± 0.5184
	EEGGAN-SepH w/o Learnable Filter	0.7840	16.5436	0.0136	0.1610 ± 0.0150	0.0019	0.0367	0.0137	0.1785 ± 0.0449
	EEGGAN + Learnable Filter	0.6578	20.0568	0.0267	0.1833 ± 0.0254	0.0032	0.0629	0.0216	0.1102 ± 0.0111
Lose	EEGGAN-SepH	0.2150	26.6978	0.0134	0.0787 ± 0.0132	0.0008	0.1004	0.0201	0.0870 ± 0.0298
	EEGGAN-SepH w/o Group Norm	1.3858	40.5554	0.0179	0.2891 ± 0.0631	0.0051	0.1202	0.0287	1.6150 ± 0.4988
	EEGGAN-SepH w/o Learnable Filter	0.7796	16.2859	0.0143	0.1554 ± 0.0176	0.0017	0.0340	0.0168	0.2031 ± 0.0521
	EEGGAN + Learnable Filter	0.6577	20.1953	0.0285	0.1829 ± 0.0232	0.0092	0.0713	0.0421	0.1736 ± 0.0131
Test-based comparisons									
Win	EEGGAN-SepH	0.1677	18.8917	0.0102	0.0569 ± 0.0084	0.0018	0.0923	0.0207	0.0666 ± 0.0211
	EEGGAN-SepH w/o Group Norm	1.4088	35.9226	0.0158	0.2883 ± 0.0702	0.0073	0.1202	0.0304	1.6301 ± 0.5209
	EEGGAN-SepH w/o Learnable Filter	0.8154	9.7178	0.0111	0.1564 ± 0.0210	0.0027	0.0378	0.0169	0.1943 ± 0.0533
	EEGGAN + Learnable Filter	0.6090	14.3928	0.0231	0.1666 ± 0.0244	0.0044	0.0665	0.0217	0.1237 ± 0.0124
Lose	EEGGAN-SepH	0.1892	17.5455	0.0103	0.0642 ± 0.0094	0.0015	0.0896	0.0211	0.1077 ± 0.0307
	EEGGAN-SepH w/o Group Norm	1.3542	35.0543	0.0165	0.2753 ± 0.0615	0.0055	0.1211	0.0340	1.4838 ± 0.4618
	EEGGAN-SepH w/o Learnable Filter	0.8369	11.8547	0.0119	0.1613 ± 0.0210	0.0029	0.0380	0.0207	0.2354 ± 0.0550
	EEGGAN + Learnable Filter	0.6317	16.7243	0.0253	0.1750 ± 0.0241	0.0111	0.0775	0.0403	0.1887 ± 0.0142

EEGGAN-SepH w/o Group Norm uses Batch Normalization instead of Group Normalization; *EEGGAN-SepH w/o Learnable Filter* removes the learnable temporal smoothing filter; *EEGGAN + Learnable Filter* adds the learnable filter alone to the

original EEGGAN output head. Everything else is unchanged. FD is computed on xDAWN, Inception-v3, REVE and EEGNet features; MMD on xDAWN, CLIP, REVE and EEGNet features. Bold values indicate the lowest score for each metric, separately within each class and within the Train-based and Test-based comparisons. EEGNet-based scores are mean $\pm$ std over 3 random seeds of the feature extractor; other scores refer to a single run.

Table 5: Ablation of the cross-electrode mixing on BCIAUT-P300 Dataset.

		FD ↓				MMD ↓			
Class	Comparison	xDAWN	Inception	REVE	EEGNet	xDAWN	CLIP	REVE	EEGNet
Train-based comparisons									
Non-Target	EEGGAN-SepH	3.9937	1.2364	0.0140	0.0489 ± 0.0013	0.0209	0.0011	0.0036	0.0500 ± 0.0041
	EEGGAN-SepH-CW	3.8492	1.2772	0.0137	0.0471 ± 0.0004	0.0303	0.0011	0.0033	0.0343 ± 0.0051
Target	EEGGAN-SepH	3.5209	1.0875	0.0144	0.0411 ± 0.0015	0.0190	0.0004	0.0030	0.0404 ± 0.0060
	EEGGAN-SepH-CW	3.5961	1.1532	0.0143	0.0422 ± 0.0016	0.0305	0.0007	0.0027	0.0314 ± 0.0047
Test-based comparisons									
Non-Target	EEGGAN-SepH	4.3598	1.4075	0.0144	0.0504 ± 0.0019	0.0170	0.0018	0.0033	0.0460 ± 0.0080
	EEGGAN-SepH-CW	4.1364	1.3563	0.0141	0.0478 ± 0.0008	0.0249	0.0018	0.0038	0.0331 ± 0.0090
Target	EEGGAN-SepH	3.9391	1.3189	0.0147	0.0461 ± 0.0013	0.0287	0.0012	0.0041	0.0625 ± 0.0046
	EEGGAN-SepH-CW	4.2042	1.2475	0.0148	0.0468 ± 0.0010	0.0409	0.0015	0.0041	0.0458 ± 0.0023

EEGGAN-SepH-CW replaces the pointwise (1×1) convolution of *EEGGAN-SepH* with a channel-wise one (groups = C): each output electrode is then a weighted sum of its own D intermediate maps, with no interaction between electrodes at this stage (16 instead of 128 parameters). Everything else is unchanged. FD is computed on xDAWN, Inception-v3, REVE and EEGNet features; MMD on xDAWN, CLIP, REVE and EEGNet features. Bold values indicate the lowest score for each metric, separately within each class and within the Train-based and Test-based comparisons. EEGNet-based scores are mean $\pm$ std over 3 random seeds of the feature extractor; other scores refer to a single run.

6 Discussion

6.1 Advantages of the Proposed Work

The limited availability of annotated EEG data underlines the practical relevance of this work. The results indicate that the signals generated by EEGGAN-SepH match real recordings more than the EEGGAN, supporting the use of the proposed model as a tool to overcome data-privacy

constraints, to support the training of foundation models, to mitigate inter-subject, inter-class, and inter-task variability, and to provide extended datasets for the analyses of clinicians, engineers, and AI practitioners.

The proposed architecture overcomes the limitations of the original EEGGAN through the SepH module, representing more than just a gaussian filter post generation. This design preserves EEG features, yielding synthetic EEG that is plausible both in terms of feature distributions and

of individual characteristics, as reflected by the FD and MMD scores obtained on two distinct datasets. The ablation study further highlights the contribution of two additional components of the module: Group Normalization, which stabilizes training compared to Batch Normalization, and the learnable temporal smoothing filter, which suppresses the high-frequency noise introduced during generation. Conversely, the auxiliary autoencoder degraded performance on Dataset 1, particularly in terms of quantitative metrics and most notably for EEGGAN-SepH, whose output head is designed to output EEG signals directly, without any decompression stage. For this reason, the autoencoder was not employed in the Dataset 2 experiments.

EEGGAN-SepH generates synthetic data that likely cover regions of the feature space not represented in the limited fine-tuning data, thereby enriching the learned representation and bringing performance closer to that obtained by pre-training on real data, which mitigates the problem of limited data.

With regard to evaluation, the gaussian noise injection test revealed a relevant limitation of FID on time-frequency images, which fails to detect mild noise added to EEG signals. Since state-of-the-art works commonly compare generative models on the basis of this metric, their conclusions may overlook differences in noise content that are critical in the EEG domain. The proposed

evaluation pipeline addresses this issue: FD and MMD computed on features from EEG-specific extractors proved more sensitive and informative, and should therefore be regarded as the primary fidelity metrics for synthetic EEG, with FID adopted only as a complementary measure.

6.2 Limitations

Although EEGGAN-SepH proves highly reliable in qualitative terms, the results reveal that noticeable differences from real data persist in terms of fidelity and downstream classification performance. This gap could potentially be narrowed by adopting alternative architectures reviewed in the state of the art, such as diffusion models; however, their sensitivity to small and imbalanced training sets, together with the iterative denoising process that makes them unsuitable for real-time BCI applications, discourages their use in this context.

Generators are trained on the calibration data of all subjects and sessions, including the session used for fine-tuning; only synthetic samples conditioned on that session are excluded from the pre-training pool. The generator may thus encode information from calibration trials beyond the fine-tuning budget, which could favour synthetic pre-training in the low-data regimes.

A further limitation concerns the pointwise convolution, which does not bring a measurable

improvement in the spatial representation. Overall, the SepH module adds value on the temporal and frequency content of the generated signals through the separable convolution and the smoothing filter, whereas the spatial information appears to be already well captured by EEGGAN itself: the spatial projection and the pointwise convolution redistribute the work of the whole architecture across electrodes, but do not yield more faithful spatial features.

7 Conclusion

This work addresses two gaps in synthetic EEG generation: generative networks lack an EEG-oriented module, and fidelity metrics that overlook the physiological nature of the signal. We proposed a GAN-based generative model that synthesizes large amounts of EEG data from noise, more closely approximating the statistical distribution of the signal features and its physiological morphology. This is achieved through the SepH, applied as the final stage of the generator to map the transformer output into a synthetic EEG signal. Furthermore, we evaluated the synthetic data with metrics that assess the fidelity of the physiological features of the EEG signal, mitigating the domain shift introduced by feature extractors pretrained on other domains.

The evaluation was carried out on two datasets. On the Reinforcement Learning dataset,

the FD computed on EEGNet features between the test set and the data generated by the proposed model is 0.0569 and 0.0642 for the Win and Lose classes, respectively, compared to 0.1297 and 0.1351 for EEGGAN; the corresponding MMD values are 0.0666 and 0.1077, compared to 0.0992 and 0.1197. On BCIAUT-P300, the FD on EEGNet features is 0.0504 and 0.0461 for the Non-Target and Target classes, compared to 0.0682 and 0.0619 for EEGGAN, while the MMD on the same features is 0.0460 and 0.0625, compared to 0.0854 and 0.1101. EEGGAN-SepH also proved effective in terms of utility metrics: with a limited calibration set of 25 targets and 175 non-targets, pre-training on its generated data increased the balanced accuracy from 62.58% to 68.66%.

Removing the auxiliary autoencoder proved beneficial in terms of both qualitative and fidelity metrics, particularly for EEGGAN-SepH, whose output head is designed to return uncompressed EEG signals. The evaluation pipeline was validated by injecting gaussian noise into real data from Dataset 2, and the metrics responded consistently to increasing noise levels. FD and MMD computed on EEGNet and REVE features proved to be the most effective in this domain, outperforming the standard FID. Finally, the ablation study confirmed the benefit of the two modifications introduced with respect to the original EEGNet block: the learnable temporal smoothing filter at the end of the output head and

the use of group normalization instead of batch normalization.

In future work, we aim to develop a diffusion model that combines fast generation with suitability for limited datasets, and to compare its performance with that of the proposed framework. To assess the spatial capability of EEGGAN-SepH, we aim to train it on a dataset with 32 or 64 channels.

8 Declarations

Availability of data and materials. The source code supporting the findings of this study is available in the EEGGAN-SepH GitHub repository, <https://github.com/GiorgioSV99/EEGGAN-SepH>. The Reinforcement Learning dataset is available in the same repository. The BCIAUT-P300 dataset is available in the Kaggle repository, <https://www.kaggle.com/datasets/disbeat/bciaut-p300>. The synthetic data generated during the current study are not deposited in a public repository, but they can be regenerated using the pretrained models provided in the EEGGAN-SepH repository.

Competing interests. The authors declare no financial or non-financial competing interests.

Funding. Not applicable.

Authors' contributions. GSV conceived the study, developed the method, performed the

experiments and wrote the original draft. AF, MRM, MC and GF supervised the work and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate.

This study is a secondary analysis of two publicly available EEG datasets; no new data were collected. Ethical approval and informed consent were secured by the original investigators.

Acknowledgements. Not applicable.

References

- [1] Kirschstein, T., Köhling, R.: What is the source of the EEG? Clin EEG Neurosci **40**(3), 146–149 (2009)
- [2] Friedman, D., Claassen, J., Hirsch, L.J.: Continuous electroencephalogram monitoring in the intensive care unit. Anesth Analg **109**(2), 506–523 (2009)
- [3] Roohi-Azizi, M., Azimi, L., Heysieattalab, S., Aamidfar, M.: Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders. Med J Islam Repub Iran **31**, 53 (2017)
- [4] Helfrich, R.F., Knight, R.T.: Cognitive neurophysiology: Event-related potentials. Handb Clin Neurol **160**, 543–558 (2019)

- [5] Rezaee, K., Ghayoumi Zadeh, H., Fayazi, A.: Parkinson's disease classification using optimized attention-based deep learning from EEG signals with interpretable sub-band topography. *Brain Inform* **13**(1), 29 (2026) <https://doi.org/10.1186/s40708-026-00317-x>
- [6] Qu, Y., Lou, X., Meng, H., Li, Z., Wang, J., Mao, K., Li, X.: EEG-DBNet: a dual-branch framework for temporal-spectral representation learning of motor imagery electroencephalography. *Brain Inform* **13**(1), 33 (2026) <https://doi.org/10.1186/s40708-026-00321-1>
- [7] Scavalli Veccia, G., Mehta, N., Bendor, S., Grover, P., Vourvopoulos, A., Figueiredo, P.: Motor imagery EEG classification with convolutional neural networks: Impact of reduced trials, channels and samples across different datasets. In: 2025 33rd European Signal Processing Conference (EUSIPCO), pp. 1487–1491 (2025). <https://doi.org/10.23919/EUSIPCO63237.2025.11226481>
- [8] Barile, B., Marzullo, A., Stamile, C., Durand-Dubief, F., Sappey-Marinier, D.: Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. *Comput Methods Programs Biomed* **206**, 106113 (2021)
- [9] Li, C., Wei, Y., Chen, X., Schönlieb, C.-B.: BrainNetGAN: Data augmentation of brain connectivity using generative adversarial network for dementia classification. In: MICCAI Workshop on Deep Generative Models, pp. 103–111 (2021)
- [10] Yin, K., Lim, E.Y., Lee, S.-W.: GITGAN: Generative inter-subject transfer for EEG motor imagery analysis. *Pattern Recognit* **146**, 110015 (2024) <https://doi.org/10.1016/j.patcog.2023.110015>
- [11] Luo, Y., Zhu, L.-Z., Wan, Z.-Y., Lu, B.-L.: Data augmentation for enhancing EEG-based emotion recognition with deep generative models. *arXiv preprint arXiv:2006.05331* (2020) <https://doi.org/10.48550/arXiv.2006.05331>
- [12] Kumar, A., Soni, S., Chauhan, S., Kaur, S., Sharma, R., Kalsi, P., Chauhan, R., Birla, A.: Navigating the realm of generative models: GANs, diffusion, limitations, and future prospects—a review. In: Tanwar, S., Singh, P.K., Ganzha, M., Epiphaniou, G. (eds.) *Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security*, pp. 301–319. Springer, Singapore (2024). https://doi.org/10.1007/978-981-97-2550-2_23

- [13] Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., Kumar, S.: Rethinking FID: Towards a better evaluation metric for image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9307–9315 (2024)
- [14] Mahey, P., Toussi, N., Purnomu, G., Herdman, A.T.: Generative adversarial network (GAN) for simulating electroencephalography. *Brain Topogr* **36**(5), 661–670 (2023)
- [15] Panwar, S., Rad, P., Jung, T.-P., Huang, Y.: Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events. *IEEE Trans Neural Syst Rehabil Eng* **28**(8), 1720–1730 (2020)
- [16] Torma, S., Szegletes, L.: Generative modeling and augmentation of EEG signals using improved diffusion probabilistic models. *J Neural Eng* **22**(1), 016001 (2025)
- [17] Williams, C., Weinhardt, D., Hewson, J., Plomecka, M.B., Langer, N., Musslick, S.: EEG-GAN: A generative EEG augmentation toolkit for enhancing neural classification. *bioRxiv* (2025) <https://doi.org/10.1101/2025.06.23.661164>
- [18] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. *J Mach Learn Res* **13**(25), 723–773 (2012)
- [19] Abdelfattah, S.M., Abdelrahman, G.M., Wang, M.: Augmenting the size of EEG datasets using generative adversarial networks. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–6 (2018). <https://doi.org/10.1109/IJCNN.2018.8489727>
- [20] Fan, J., Sun, C., Chen, C., Jiang, X., Liu, X., Zhao, X., Meng, L., Dai, C., Chen, W.: EEG data augmentation: towards class imbalance problem in sleep staging tasks. *J Neural Eng* **17**(5), 056017 (2020)
- [21] Osman, Y.B.M., Elsanosi, A.H.M., Jing, C., Shen, Y., Kuai, H., Wang, S., Ng, M.K.-P., Lei, B., Wang, S.: Diffusion models for brain imaging computing: a survey of frameworks and applications. *Brain Inform* **13**(1), 19 (2026) <https://doi.org/10.1186/s40708-026-00301-5>
- [22] Carrle, F.P., Hollenbenders, Y., Reichenbach, A.: Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depressive disorder. *Front Neurosci* **17**, 1219133 (2023) <https://doi.org/10.3389/fnins.2023.1219133>

- [23] Rommel, C., Paillard, J., Moreau, T., Gramfort, A.: Data augmentation for learning predictive models on EEG: a systematic comparison. *J Neural Eng* **19**(6), 066020 (2022)
- [24] Rasheed, K., Qadir, J., O'Brien, T.J., Kuhlmann, L., Razi, A.: A generative model to synthesize EEG data for epileptic seizure prediction. *IEEE Trans Neural Syst Rehabil Eng* **29**, 2322–2332 (2021)
- [25] Jiang, W., Zhao, L., Lu, B.-l.: Large brain model for learning generic representations with tremendous EEG data in BCI. In: International Conference on Learning Representations (ICLR) (2024). <https://openreview.net/forum?id=QzTpTRVtrP>
- [26] Wang, Z., He, Z., He, X., Wang, H., Jia, T., Luo, J., Li, S., Chen, X., Wu, D.: Synthetic data generation for brain-computer interfaces: Overview, benchmarking, and future directions. *arXiv preprint arXiv:2603.12296* (2026) <https://doi.org/10.48550/arXiv.2603.12296>
- [27] Couchman, J., Kaparounakis, O., Samarakoon, C., Stanley-Marbell, P.: Simulated eyeblink artifact removal with ICA: Effect of measurement uncertainty. *arXiv preprint arXiv:2410.03261* (2024) <https://doi.org/10.48550/arXiv.2410.03261>
- [28] Frässle, S., Yao, Y., Schöbi, D., Aponte, E.A., Heinzle, J., Stephan, K.E.: Generative models for clinical applications in computational psychiatry. *Wiley Interdiscip Rev Cogn Sci* **9**(3), 1460 (2018)
- [29] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations (ICLR) (2018). <https://arxiv.org/abs/1710.09412>
- [30] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* **16**, 321–357 (2002) <https://doi.org/10.1613/jair.953>
- [31] Wang, F., Zhong, S.-h., Peng, J., Jiang, J., Liu, Y.: Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In: International Conference on Multimedia Modeling (MMM), pp. 82–93 (2018)
- [32] Cheng, J.Y., Goh, H., Dogrusoz, K., Tuzel, O., Azemi, E.: Subject-aware contrastive learning for biosignals. *arXiv preprint arXiv:2007.04871* (2020) <https://doi.org/10.48550/arXiv.2007.04871>

- [33] Saeed, A., Grangier, D., Pietquin, O., Zeghidour, N.: Learning from heterogeneous EEG signals with differentiable channel reordering. arXiv preprint arXiv:2010.13694 (2020) <https://doi.org/10.48550/arXiv.2010.13694>
- [34] Schwabedal, J.T., Snyder, J.C., Cakmak, A., Nemati, S., Clifford, G.D.: Addressing class imbalance in classification problems of noisy signals by using Fourier transform surrogates. arXiv preprint arXiv:1806.08675 (2018) <https://doi.org/10.48550/arXiv.1806.08675>
- [35] Lashgari, E., Liang, D., Maoz, U.: Data augmentation for deep-learning-based electroencephalography. *J Neurosci Methods* **346**, 108885 (2020) <https://doi.org/10.1016/j.jneumeth.2020.108885>
- [36] Oussidi, A., Elhassouny, A.: Deep generative models: Survey. In: 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1–8 (2018). <https://doi.org/10.1109/ISACV.2018.8354080>
- [37] Krichen, M.: Generative adversarial networks. In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7 (2023)
- [38] Weng, L.: From GAN to WGAN. arXiv preprint arXiv:1904.08994 (2019) <https://doi.org/10.48550/arXiv.1904.08994>
- [39] Han, F., Liang, Y., Ling, Q., Han, H.: An improved conditional Wasserstein GAN with gradient penalty for gene expression profiling data augmentation based on data segmentation and depth feature constraint. *IEEE Trans Comput Biol Bioinform* **22**(4), 1401–1414 (2025)
- [40] Reynolds, D.A.: Gaussian mixture models. In: Li, S.Z., Jain, A. (eds.) *Encyclopedia of Biometrics*, pp. 659–663. Springer, Boston, MA (2009). https://doi.org/10.1007/978-0-387-73003-5_196
- [41] Xu, M., Chen, Y., Wang, Y., Wang, D., Liu, Z., Zhang, L.: BWGAN-GP: An EEG data generation method for class imbalance problem in RSVP tasks. *IEEE Trans Neural Syst Rehabil Eng* **30**, 251–263 (2022) <https://doi.org/10.1109/TNSRE.2022.3145515>
- [42] Li, X., Metsis, V., Wang, H., Ngu, A.H.H.: TTS-GAN: A transformer-based time-series generative adversarial network. In: *International Conference on Artificial Intelligence in Medicine (AIME)*, pp. 133–143 (2022)
- [43] Luo, T.-j., Cai, Z.: Diffusion models-based motor imagery EEG sample augmentation via mixup strategy. *Expert Syst Appl* **262**, 125585 (2025)

- [44] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: International Conference on Learning Representations (ICLR) (2016)
- [45] Zhang, J., Zhang, H., Yang, Y.: Generative diffusion meets domain adaptation: a framework for EEG cross-subject motor imagery classification. *Brain Inform* **13**(1), 24 (2026) <https://doi.org/10.1186/s40708-026-00308-y>
- [46] Abul-Fazl, S., Dina, R., Fairuza, H.: Diffusion models at scale: Techniques, applications, and challenges. *Preprints* (2025) <https://doi.org/10.20944/preprints202502.0029.v1>
- [47] Ibrahim, M., Al Khalil, Y., Amirrajab, S., Sun, C., Breeuwer, M., Pluim, J., Elen, B., Ertaylan, G., Dumontier, M.: Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. *Comput Biol Med* **189**, 109834 (2025) <https://doi.org/10.1016/j.compbiomed.2025.109834>
- [48] He, C., Liu, J., Zhu, Y., Du, W.: Data augmentation for deep neural networks model in EEG classification task: A review. *Front Hum Neurosci* **15**, 765525 (2021) <https://doi.org/10.3389/fnhum.2021.765525>
- [49] Hazra, D., Byun, Y.-C.: SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation. *Biology (Basel)* **9**(12), 441 (2020) <https://doi.org/10.3390/biology9120441>
- [50] Wang, H., Chen, X., Yang, Y., Zhou, K., Lv, M., Wang, D., Zhang, W.: EEG signal denoising using pix2pix GAN: enhancing neurological data analysis. *arXiv preprint arXiv:2411.13288* (2024) <https://doi.org/10.48550/arXiv.2411.13288>
- [51] Brophy, E., Redmond, P., Fleury, A., De Vos, M., Boylan, G., Ward, T.: Denoising EEG signals for real-world BCI applications using GANs. *Front Neuroergon* **2**, 805573 (2022)
- [52] Vo, K., Vishwanath, M., Srinivasan, R., Dutt, N., Cao, H.: Composing graphical models with generative adversarial networks for EEG signal modeling. In: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1231–1235 (2022)
- [53] Polychronaki, G.E., Ktonas, P.Y., Gatzonis, S., Siatouni, A., Asvestas, P.A., Tsekou, H., Sakas, D., Nikita, K.S.: Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. *J Neural Eng* **7**(4), 046007 (2010)
- [54] Esteller, R., Vachtsevanos, G., Echauz, J.,

- Litt, B.: A comparison of fractal dimension algorithms using synthetic and experimental data. In: 1999 IEEE International Symposium on Circuits and Systems (ISCAS), vol. 3, pp. 199–202 (1999)
- [55] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Advances in Neural Information Processing Systems, vol. 30, pp. 6626–6637 (2017)
- [56] Hartmann, K.G., Schirrmeister, R.T., Ball, T.: EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals. arXiv preprint arXiv:1806.01875 (2018) <https://doi.org/10.48550/arXiv.1806.01875>
- [57] Aznan, N.K.N., Atapour-Abarghouei, A., Bonner, S., Connolly, J.D., Al Moubayed, N., Breckon, T.P.: Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved SSVEP classification. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2019)
- [58] Cinquetti, E., Menegaz, G., Storti, S.F.: Toward in-silico data assessment for passive BCIs: generating EEG rhythms with GANs.
- J Neural Eng **22**(6), 066007 (2025)
- [59] Vos, G., Ebrahimpour, M., Eijk, L., Sarnyai, Z., Azghadi, M.R.: A statistical approach for synthetic EEG data generation. arXiv preprint arXiv:2504.16143 (2025) <https://doi.org/10.48550/arXiv.2504.16143>
- [60] El Emam, K., Mosquera, L., Fang, X., El-Hussuna, A.: Utility metrics for evaluating synthetic health data generation methods: Validation study. JMIR Med Inform **10**(4), 35734 (2022) <https://doi.org/10.2196/35734>
- [61] Figueira, A., Vaz, B.: Survey on synthetic data generation, evaluation methods and GANs. Mathematics **10**(15), 2733 (2022) <https://doi.org/10.3390/math10152733>
- [62] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEG-Net: a compact convolutional neural network for EEG-based brain–computer interfaces. J Neural Eng **15**(5), 056013 (2018) <https://doi.org/10.1088/1741-2552/aace8c>
- [63] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems, vol. 30, pp. 5767–5777 (2017)
- [64] Rivet, B., Souloumiac, A., Attina, V., Gilbert, G.: xDAWN algorithm to enhance evoked

- potentials: application to brain–computer interface. *IEEE Trans Biomed Eng* **56**(8), 2035–2043 (2009) <https://doi.org/10.1109/TBME.2009.2012869>
- [65] El Ouahidi, Y., Lys, J., Thölke, P., Farrugia, N., Padeloup, B., Gripon, V., Jerbi, K., Lioi, G.: REVE: A foundation model for EEG – adapting to any setup with large-scale pre-training on 25,000 subjects. In: *Advances in Neural Information Processing Systems*, vol. 38, pp. 22541–22577 (2025). <https://doi.org/10.52202/085713-0760>
- [66] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826 (2016). <https://doi.org/10.1109/CVPR.2016.308>
- [67] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: *Proceedings of the 38th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research*, vol. 139, pp. 8748–8763 (2021)
- [68] Williams, C.C., Ferguson, T.D., Hassall, C.D., Abimbola, W., Krigolson, O.E.: The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology* **58**(2), 13722 (2021) <https://doi.org/10.1111/psyp.13722>
- [69] Simões, M., Borra, D., Santamaría-Vázquez, E., Bittencourt-Villalpando, M., Krzemiński, D., Miladinović, A., Schmid, T., Zhao, H., Amaral, C., Direito, B., Henriques, J., Carvalho, P., Castelo-Branco, M.: BCIAUT-P300: A multi-session and multi-subject benchmark dataset on autism for P300-based brain-computer-interfaces. *Front Neurosci* **14**, 568104 (2020) <https://doi.org/10.3389/fnins.2020.568104>
- [70] Williams, C.C., Weinhardt, D., Wirzberger, M., Musslick, S.: Augmenting EEG with generative adversarial networks enhances brain decoding across classifiers and sample sizes. In: Goldwater, M., Anggoro, F.K., Hayes, B.K., Ong, D.C. (eds.) *Proceedings of the 45th Annual Conference of the Cognitive Science Society (CogSci)*, vol. 45. Cognitive Science Society, Sydney, Australia (2023). <https://escholarship.org/uc/item/9gz8g908>