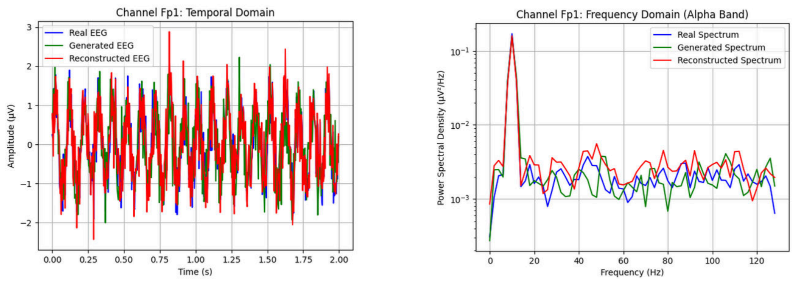
DOI: 10.3217/978-3-99161-050-2-094

# Manifold-Based Diffusion Models for Generating Synthetic Motor Imagery EEG

**Imad Eddine Tibermacine**[1*], **Christian Napoli**[1,2,3]
[1]*Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy;*
[2]*Institute for Systems Analysis and Computer Science, Italian National Research Council, Rome, Italy;*
[3]*Department of Computational Intelligence, Czestochowa University of Technology, Czestochowa, Poland*
* E-mail: tibermacine@diag.uniroma1.it

*Introduction:* Generating realistic synthetic EEG can alleviate dataset size, privacy, and session variability concerns in motor imagery (MI) brain–computer interfaces. Conventional diffusion-based models risk off-manifold artifacts, ignoring the manifold structure of EEG in space and time[1]. We employ *Symmetric Positive Definite* (SPD) matrices derived from EEG covariance to capture electrode correlations. Constraining diffusion on this SPD manifold yields higher-fidelity reconstructions and generated signals for MI tasks[2].




(a) Time-domain comparison of Real (blue), Generated (green), and Reconstructed (red) EEG signals for Channel Fp1.

(b) Spectral comparison of Real (blue), Generated (green), and Reconstructed (red) EEG signals, emphasizing alpha-band retention.

Figure 1: Combined time-domain and spectral comparisons of real, generated, and reconstructed EEG signals for Channel Fp1.

*Material, Methods, and Results:* We used the BCI Competition IV dataset 2a (BCIC-IV-2a), which includes four-class motor imagery EEG from nine subjects, with 288 trials per session recorded over 22 electrodes at 250 Hz. Preprocessing steps involved noise reduction, down-sampling to 128 Hz, and segmenting 3.5-second epochs, producing $22 \times 22$ covariance matrices on the SPD manifold. A forward stochastic differential equation (SDE) was defined,

$$d\mathbf{z} = f(\mathbf{z}, t)\, dt + g(\mathbf{z}, t)\, d\mathbf{w},$$

where $\mathbf{z}$ represents latent SPD coordinates, and $f(\cdot), g(\cdot)$ regulate manifold-constrained noise injection. The reverse-time SDE was approximated by a neural network using mean squared error and cosine similarity, incorporating *log* and *exp* mapping operations alongside a multi-scale diffusion strategy. Synthetic SPD outputs were mapped back to time-domain EEG through Cholesky factorization. Figures 1a and 1b demonstrate that the generated signals preserve amplitude fluctuations and alpha-band peaks (10–12 Hz), improving classification accuracy and reducing Fréchet distance compared to Euclidean baselines.

*Discussion and Significance:* Constraining diffusion to the SPD manifold minimizes off-manifold artifacts and preserves critical temporal and spectral features of the MI EEG. The overlap between real and synthetic signals demonstrates the retention of inter-electrode correlations and alpha oscillations, making this approach valuable for data augmentation in small-sample or privacy-restricted scenarios. Future work could refine Riemannian metrics, extend to multi-class MI tasks, and validate on larger clinical datasets, solidifying manifold-based diffusion as a reliable method for generating realistic EEG data.

*References:*

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *Advances in Neural Information Processing Systems*, 2020.
[2] Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456 (2020).*



Published by Verlag der Technischen Universität Graz