Synthetic Sleep EEG Signal Generation using Latent Diffusion Models

Bruno Aristimunha, Raphael Yokoingawa de Camargo, Sylvain Chevallier, Adam G. Thomas, Oeslle Lucena, Jorge Cardoso, Walter Hugo Lopez Pinaya, Jessica Dafflon

Synthetic Sleep EEG Signal Generation using Latent Diffusion Models

Bruno Aristimunha
Université Paris-Saclay, Inria TAU, France
Federal University of ABC, Brazil
b.aristimunha@gmail.com

Raphael Y. de Camargo
Federal University of ABC,
Santo Andre, Brazil

Sylvain Chevallier
Université Paris-Saclay,
Inria TAU, France

Adam G. Thomas
Data Science and Sharing Team,
NIMH, USA

Oeslle Lucena
King's College London,
London, United Kingdom

M. Jorge Cardoso
King's College London,
London, United Kingdom

Walter H. Lopez Pinaya
King's College London,
London, United Kingdom

Jessica Dafflon
Data Science and Sharing Team, NIMH, USA
Machine Learning Team, NIMH, USA

Abstract

Electroencephalography (EEG) is a non-invasive method that allows for recording rich temporal information and is a valuable tool for diagnosing various neurological and psychiatric conditions. One of the main limitations of EEG is the low signal-to-noise ratio and the lack of data availability to train large data-hungry neural networks. Sharing large healthcare datasets is crucial to advancing medical imaging research, but privacy concerns often impede such efforts. Deep generative models have gained attention as a way to circumvent data-sharing limitations and as a possible way to generate data to improve the performance of these models. This work investigates latent diffusion models with spectral loss as deep generative modeling to generate 30-second windows of synthetic EEG signals of sleep stages. The spectral loss is essential to guarantee that the generated signal contains structured oscillations on specific frequency bands that are typical of EEG signals. We trained our models using two large sleep datasets (*Sleep EDFx* and *SHHS*) and used the Multi-Scale Structural Similarity Metric, Frechet inception distance, and a spectrogram analysis to evaluate the quality of synthetic signals. We demonstrate that the latent diffusion model can generate realistic signals with the correct neural oscillation and could, therefore, be used to overcome the scarcity of EEG data.

1 Introduction

Sleep is a fundamental cognitive task in which quality and duration are essential for human well-being. Nonetheless, many still experience sleep disorders, such as insomnia, sleep apnea, and bruxism, linked to physical and emotional conditions [1, 2, 3, 4, 5, 6, 7]. Polysomnography (PSG) is the standard test to capture various biosignals throughout the night and monitor sleep. PSG involves categorizing 30-second time intervals into different sleep stages and using the distribution of

1st Workshop on Deep Generative Models for Health at NeurIPS 2023.

these stages for diagnosing sleep disorders. Electroencephalography (EEG) data is considered the most reliable PSG biomarker across all sleep stages [8, 9]. However, the variability of sleep data poses significant challenges, including individual brain signal signature, low signal-to-noise ratio, age-related variability, pathologies, and cognitive state on the distribution of sleep stages [9].

Deep generative methods, especially diffusion models, have reached state-of-the-art performance in many medical imaging tasks, such as detecting anomalies, segmenting tasks, generating synthetic data, and improving the traditional supervision classification [10, 11, 12, 13, 14, 15, 16, 17]. However, deep generative applied to bio-signal is still an active area of research due to the challenges of EEG decoding. Generative adversarial networks (GANs) have been used to generate EEG data [18, 19, 20]. However, the use of GANs poses challenges, including difficulties in scaling, the well-known issue of mode collapsing, and a lack of flexibility in learning new tasks, even within the same domain [21, 22, 23]. On the other hand, studies involving diffusion models in EEG have concentrated on Brain-Computer Interface (BCI) applications and/or have been limited to datasets with fewer than 30 individuals [24, 25, 26, 27]. Diffusion models have a strong theoretical foundation and cannot only capture the dynamics but also more context of signal structures when compared with GANs [16, 17, 28, 29]. Nonetheless, Diffusion models can extract essential factors that affect the data, such as age, pathology, and task, from experiments or other factors that impact EEG recordings.

In this work, we trained a Latent Diffusion Model with polysomnogram data to generate synthetic EEG signals with a 30-second window. We conduct experiments using two large sleep datasets: Sleep EDFx [30] and Sleep Heart Health Study (SHHS) [31]. We describe the model and methodology in more detail in Appendix A. Our model was trained to work without restriction on brain electrode position. We used an AutoEncoder with KL regularization (AE-KL) [17] to compress the EEG signal and relied on the obtained latent space to train the Latent Diffusion Model (LDM). Then, we evaluate the quality of the generated synthetic signal by implementing well-established generative metrics: the Fréchet inception distance (FID), Multi-Scale Structural Similarity Metric (MS-SSIM), and power spectrum analysis within the sleep band. Our model can generate EEG trials that closely approximate the significant brainwave interval associated with fundamental keys for the sleep stages, including δ $(0.1 - 4\text{Hz})$, θ $(4 - 8\text{Hz})$, and α $(8 - 12\text{Hz})$ waves. The code and models are publicly available at `https://github.com/bruAristimunha/Synthetic-Sleep-EEG-Signal-Generation-using-Latent-Diffusion-Models`.

2 Methodology

Dataset We employed two extensive sleep stage datasets to train our deep generative model: the Physionet Sleep-EDFx dataset (*Sleep EDFx*)[30] and Sleep Heart Health Study (*SHHS*)[31]. More pre-processing and epoch details can be found in Appendix A.1.

AutoEncoder with Kullback-Leibler regularization (AE-KL) In accordance with the model proposed by [17, 13], we employed an AE-KL to compress the EEG windows.

The training of the AE-KL involved a combination of four objective loss functions: The L1 loss (ℓ_{recons}) to measure the discrepancy between the input ($\mathbf{x}$) and the reconstructed output ($\widehat{\mathbf{x}}$); a patch-based adversarial objective (ℓ_{adv}); [32], the Kullback-Leibler Divergence (ℓ_{kl}); and the Jukebox loss (ℓ_{spec}) [33, 34] to enhance spectral component learning. The Jukebox loss, denoted as $\ell_{\text{spec}}(\mathbf{x}, \widehat{\mathbf{x}})$, quantifies the L2 norm difference between the absolute values of Short-Time Fourier Transforms (STFT) of the input ($\mathbf{x}$) and the reconstructed output ($\widehat{\mathbf{x}}$). We assessed its impact on signal generation through ablation studies training models with and without this loss component. See Equation 1 for an overview of all training losses in the AE-KL model.

$$\min \sum \ell_{\text{recons}}(\mathbf{x}, \widehat{\mathbf{x}}) + \ell_{\text{adv}}(\mathbf{x}, \widehat{\mathbf{x}}) + \ell_{\text{kl}}(\mathbf{z}_\mu, \mathbf{z}_\sigma) + \ell_{\text{spec}}(\mathbf{x}_i, \widehat{\mathbf{x}}), \tag{1}$$

The spectral loss is vital because sleep patterns are influenced by structured oscillations in specific frequency bands (δ, θ, α) within the brain [35]. Spectrum analysis is particularly adept at capturing and translating these dynamics from time series data into the frequency domain [35, 36]. The frequency domain augmentations, associated with spectral information, have demonstrated a remarkable improvement of the neural decoding when applied to EEG [37]. Appendix A.2.1 presents more details about the AutoEncoder model.

Latent Diffusion Model We trained a Denoising Diffusion Probabilistic Model (DDPM) to capture the distribution of a latent representation of EEG sleep signals. DDPM acts as a deep generative model that reconstructs original data points from noisy samples using a forward diffusion process. Small increments of Gaussian noise are iteratively added to a data point $\mathbf{x}_0$ drawn from the data distribution, controlled by a predefined variance schedule. As the number of steps, T, increases, the final sample $\mathbf{x}_T$ approximates a draw from an isotropic Gaussian distribution. The DDPM incorporates a reverse diffusion process that leverages graph modeling principles based on Markov chains. This reverse diffusion process facilitates the recovery of input data from noise while preserving the data's structural information. More details about the forward diffusion and reverse processes are explained in Appendix A.3.

Network training We sample 30-second windows from each electrode from Polysomnographic recording. Every 30-second window randomly sampled with overlapping was then used as a separate entry and combined with other subjects to create the batch. The first step of our model was to train an AE-KL to compress and reconstruct the EEG windows, with high-frequency components learned later during training, as shown in more detail in Appendix B.1. Once AE-KL was trained, we used the latent space as input for the LDM. Because the structural oscillation present in EEG signals is essential, we tested if adding the Jukebox loss [33, 34] would improve the oscillations of the generated signal and compared the models' performance to assess the importance of this loss. Appendix A illustrates the overview of our method.

Evaluation metrics Similar to the works from [38, 13], we used three metrics to quantitatively evaluate the fidelity of the synthetic sleep EEG: Fréchet Inception Distance (FID), Multi-Scale Structural Similarity Metric (MS-SSIM), and Power Spectral density (PSD).

The FID assesses the realism of the synthetic signals generated by calculating the distance between feature vectors calculated for real and generated signals [39]. This metric uses the W_2 Wasserstein distance to compute the difference between two probability measures. In the imaging context, this metric is called FID as it often uses the Inception model to compute the feature vectors, and since the term, FID is well established even when not using the Inception model, we will refer to this metric as FID hereafter. Our analysis used a pre-trained convolutional neural network [40, 41] trained to classify different sleep stages to extract the latent feature vectors. Lower FID scores indicate that the images are more similar, with a perfect score of zero indicating that the images are identical (Table 1).

The Multi-Scale Structural Similarity Metric (MS-SSIM) [42] evaluates the similarities between signals. MS-SSIM is an extension of SSIM that computes the structural similarity measure at multiple scales. This metric can assume values between 0 and 1, where 1 indicates perfect signal similarity. We use the MS-SSIM metric to (1) compare the AutoEncoder reconstructed signals with the real ones, where a high MS-SSIM means the latent space was able to condense all the necessary information, and (2) analyze the diversity of the synthetic images using pairs of synthetic images (Table 1). For both FID and MS-SSIM, we computed the average values from all available test samples.

In addition, we also evaluate the sleep data using a power spectral density (PSD) analysis on both the real, synthetic with spectral, and synthetic without the spectral loss data. The PSD provides a way to analyze the frequency content of a signal and understand how much power or energy is associated with each frequency component. This tool is particularly relevant for sleep problems, as specific brain rhythms are more prominent during certain sleep stages. Here, we compute the PSD by averaging PSD plots of 1000 random samples and confidence intervals based on the percentile 10% and 90%.

3 Results and Discussion

Table 1 summarizes the quantitative metrics used to evaluate the quality of the dataset. We can see that across both datasets, the MS-SSIM of the reconstruction—without (Rec) and with (Rec_{spec}) the spectral loss—and the test dataset was bigger than 0.82. This indicates that the AE-KL could learn a good latent representation of the data and reconstruct the original images with high fidelity. When comparing the LDM models trained with and without the spectral loss, we can see that the spectral models had a smaller FID and higher MS-SSIM, indicating a better result compared to the models trained without the spectral loss. Additionally, LDM_{spec}'s evaluation metrics had values closer to the real signal, suggesting that they are more similar to the real signal when compared to models trained without the spectral loss.

Table 1: Quantitative evaluation of synthetic sleep stages on different datasets. The FID and MS-SSIM were used to assess the realism of the generated and reconstructed sleep stage window. We evaluated the LDMs trained without (LDM)/with a Spectral Loss (LDM_{spec}), the metrics obtained for the Real dataset, and the MS-SSIM of the AE-KL reconstruction (Rec). $\uparrow$ / $\downarrow$ means bigger/smaller values are better.

Dataset	FID $\downarrow$			MS-SSIM $\uparrow$				
	LDM	LDM_{spec}	Real	LDM	LDM_{spec}	Real	Rec	Rec_{spec}
Sleep EDFx	11.933	**0.308**	0.015	0.205	**0.515**	0.622	**0.983**	0.907
SHHS_h	0.936	**0.168**	0.086	0.228	**0.598**	0.449	**0.969**	0.827

Figure 1 showcases the frequency content of the real in test split and synthetic data with a PSD analysis. The different sleep stages are showcased by different background colors δ $(0.5 - 4\text{Hz})$, θ $(4 - 8$ Hz), α $(8 - 12$ Hz). It is striking that in the δ range, the real and synthetic data with spectral have similar peaks and troughs, signaling that the LDM was able to learn fine structures in the signal. To our understanding, the tail decay, which is more pronounced in the synthetic data with spectral, may be associated with the filtering we performed during the pre-processing. Additionally, as indicated by [43], the alpha α interval exhibits non-linear changes during aging, with distinctions between REM and N-REM sleep stages. The rhythm variation could explain the challenges faced by the LDM in matching the alpha distribution perfectly. Notably, the PSD feature alignment of our synthetic dataset with the real dataset is a good indicator of the quality of the generative component of the LDM.

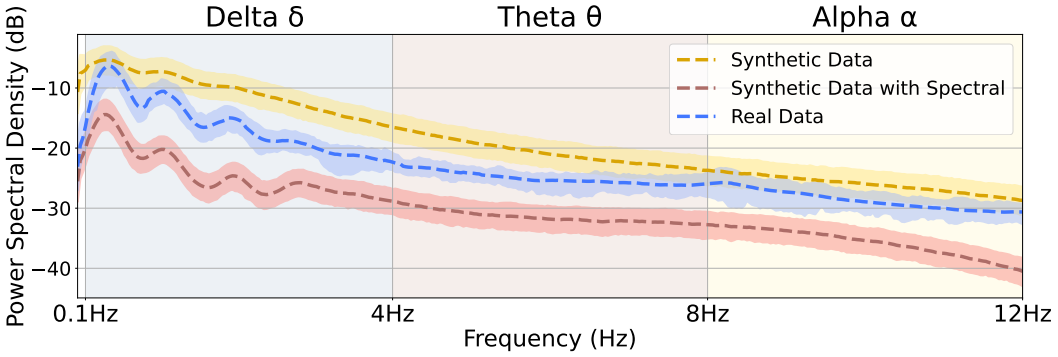

Figure 1: Averaged power spectral density of windows corresponding to the sleep data for the real in the test split and synthetic data with *Sleep EDFx* model trained with and without spectrum loss.

Analyzing the Table 1 and Figure 1, the impact of spectral loss in optimizing the AutoEncoder is clear. Given the peculiarities of brain time series, exploring the development of additional spectral loss functions is essential. This is a critical step for improving the utility of generative models in this field. A similar need is observed in audio, as documented by authors in [44], indicating substantial room for enhancing spectral loss functions.

4 Conclusions

In this work, we used latent diffusion models to create artificial EEG brain waves and generate sleep EEG signals that closely resemble existing sleep stage data. To the best of our knowledge, this is the first time diffusion models were applied to *sleep stage* data, making it a pioneering achievement. Our results demonstrate that LDMs can produce realistic sleep stage windows with sleeping patterns closely resembling the real data. Further investigation can be taken by using specific variables (i.e., age, sex, or presence/absence of pathology) to address inherent challenges posed by the imbalance present in EEG datasets. Our work highlights LDMs' promising potential in generating EEG signals and showcases the models' performance in generating sleep EEG. Another possible avenue of research is to explore the usage of diffusion models to learn directly from the raw signal instead of using a latent representation from which to learn the LDM. Utilizing the capability to produce authentic

signals featuring accurate neural oscillations could enhance the efficacy of EEG models, addressing both the shortage and imbalance inherent in EEG data.

Acknowledgments and Disclosure of Funding

We would like to thank Physionet and National Sleep Research Resource for providing the data. No ethics application was needed for the current study; ethics approval was obtained and approved by the ethics committee of the leading institution of each study.

We would like to thank Petru-Daniel Tudosiu for introducing the jukebox loss. This work was partially supported by the DATA IA (project YARN). The work of BA was supported in part by the CAPES under Grant 001 and Data IA mobilité internationale. OL was supported by King's College London. The work of WHLP MJC was supported by Wellcome Innovations under Grant [WT213038/Z/18/Z]. JD is supported by the Intramural Research Program of the NIMH (ZIC-MH002960 and ZIC-MH002968).

References

[1] W Ward Flemons. Obstructive sleep apnea. *New England Journal of Medicine*, 347(7):498–504, 2002.

[2] Marlene A Reimer and W Ward Flemons. Quality of life in sleep disorders. *Sleep medicine reviews*, 7(4):335–349, 2003.

[3] Sudhansu Chokroverty et al. Overview of sleep & sleep disorders. *Indian J Med Res*, 131(2):126–140, 2010.

[4] Kazuo Chin. Overcoming sleep disordered breathing and ensuring sufficient good sleep time for a healthy life expectancy. *Proceedings of the Japan Academy, Series B*, 93(8):609–629, 2017.

[5] Jae Wook Cho and Jeanne F Duffy. Sleep, sleep disorders, and sexual dysfunction. *The world journal of men's health*, 37(3):261–275, 2019.

[6] Bo-Huei Huang, Borja del Pozo Cruz, Armando Teixeira-Pinto, Peter A. Cistulli, and Emmanuel Stamatakis. Influence of poor sleep on cardiovascular disease-free life expectancy: a multi-resource-based population cohort study. *BMC Medicine*, 21(1):75, 2023.

[7] Saverio Stranges, William Tigbe, Francesc Xavier Gómez-Olivé, Margaret Thorogood, and Ngianga-Bakwin Kandala. Sleep problems: an emerging global epidemic? *Sleep*, 35(8):1173–1181, 2012.

[8] Brendan P Lucey, Jennifer S Mcleland, Cristina D Toedebusch, Jill Boyd, John C Morris, Eric C Landsness, Kelvin Yamada, and David M Holtzman. Comparison of a single-channel EEG sleep study to polysomnography. *J Sleep Res*, 25(6):625–635, 2016.

[9] Paul Bouchequet, Geoffroy Solelhac, Thomas Andrillon, Francesco Romano, Marco Brigham, Mounir Chennaoui, and Damien Leger. Quantifying Importance Of EEG Spectral Domain Features In Automatic Diagnosis Of Chronic Insomnia. *Sleep*, 42:A130, 2019.

[10] Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J. Fleet. Synthetic data from diffusion models improves imagenet classification, 2023.

[11] Walter Hugo Lopez Pinaya, Petru-Daniel Tudosiu, Robert Gray, Geraint Rees, Parashkev Nachev, Sébastien Ourselin, and M Jorge Cardoso. Unsupervised brain anomaly detection and segmentation with transformers. *arXiv preprint arXiv:2102.11650*, 2021.

[12] Walter HL Pinaya, Mark S Graham, Robert Gray, Pedro F Da Costa, Petru-Daniel Tudosiu, Paul Wright, Yee H Mah, Andrew D MacKinnon, James T Teo, Rolf Jager, et al. Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. *arXiv preprint arXiv:2206.03461*, 2022.

[13] Walter H. L. Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F. Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Brain Imaging Generation With Latent Diffusion Models. In *DGM4:MICCAI*, pages 117–126. Springer, 2022.

[14] Virginia Fernandez, Walter Hugo Lopez Pinaya, Pedro Borges, Petru-Daniel Tudosiu, Mark S. Graham, Tom Vercauteren, and M. Jorge Cardoso. Can segmentation models be trained with fully synthetically generated data? In *SASHIMI*, pages 79–90. Springer, 2022.

[15] Virginia Fernandez, Pedro Sanchez, Walter Hugo Lopez Pinaya, Grzegorz Jacenków, Sotirios A Tsaftaris, and Jorge Cardoso. Privacy Distillation: Reducing Re-identification Risk of Multi-modal Diffusion Models. *arXiv preprint arXiv:2306.01322*, 2023.

[16] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *NeurIPS*, 34, 2021.

[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022.

[18] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, 2019.

[19] Isaac A. Corley and Yufei Huang. Deep EEG super-resolution: Upsampling EEG spatial resolution with Generative Adversarial Networks. In *BHI*, pages 100–103, 2018.

[20] Rongkai Zhang, Ying Zeng, Li Tong, Jun Shu, Runnan Lu, Kai Yang, Zhongrui Li, and Bin Yan. ERP-WGAN: A data augmentation method for EEG single-trial detection. *Journal of Neuroscience Methods*, 376:109621, 2022.

[21] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating Diverse High-Fidelity Images with VQ-VAE-2. In *NeurIPS*, volume 32. Curran Associates, Inc., 2019.

[22] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[23] Charlie Nash, Jacob Menick, Sander Dieleman, and Peter W. Battaglia. Generating Images with Sparse Representations. 2021.

[24] Giulio Tosato, Cesare M. Dalbagno, and Francesco Fumagalli. EEG synthetic data generation using probabilistic diffusion models. *CoRR*, abs/2303.06068, 2023.

[25] Yiqun Duan, Jinzhao Zhou, Zhen Wang, Yu-Cheng Chang, Yu-Kai Wang, and Chin-Teng Lin. Domain-Specific Denoising Diffusion Probabilistic Models for Brain Dynamics. *CoRR*, abs/2305.04200, 2023.

[26] Szabolcs Torma and Luca Szegletes. EEGWave: a Denoising Diffusion Probabilistic Approach for EEG Signal Generation. EasyChair Preprint no. 10275, 2023.

[27] Kai Shu, Yuchang Zhao, Le Wu, Aiping Liu, Ruobing Qian, and Xun Chen. Data Augmentation for Seizure Prediction with Generative Diffusion Model. *CoRR*, abs/2306.08256, 2023.

[28] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents. 2022.

[29] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Raphael Gontijo-Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *NeurIPS*, 2022.

[30] Bob Kemp, Aeilko H Zwinderman, Bert Tuk, Hilbert AC Kamphuisen, and Josefien JL Oberye. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.*, 47(9):1185–1194, 2000.

[31] Stuart F Quan, Barbara V Howard, Conrad Iber, James P Kiley, F Javier Nieto, George T O'Connor, David M Rapoport, Susan Redline, John Robbins, Jonathan M Samet, et al. The sleep heart health study: design, rationale, and methods. *Sleep*, 20(12):1077–1085, 1997.

[32] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming Transformers for High-Resolution Image Synthesis. In *CVPR*, pages 12873–12883, 2021.

[33] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for music. *CoRR*, abs/2005.00341, 2020.

[34] Walter HL Pinaya, Mark S Graham, Eric Kerfoot, Petru-Daniel Tudosiu, Jessica Dafflon, Virginia Fernandez, Pedro Sanchez, Julia Wolleb, Pedro F da Costa, Ashay Patel, et al. Generative ai for medical imaging: extending the monai framework. *arXiv preprint arXiv:2307.15208*, 2023.

[35] Michael J. Prerau, Ritchie E. Brown, Matt T. Bianchi, Jeffrey M. Ellenbogen, and Patrick L. Purdon. Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology*, 32(1):60–92, 2017.

[36] SM Purcell, DS Manoach, C Demanuele, BE Cade, S Mariani, R Cox, G Panagiotaropoulou, R Saxena, JQ Pan, JW Smoller, et al. Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource. *Nature communications*, 8(1):15930, 2017.

[37] Cédric Rommel, Joseph Paillard, Thomas Moreau, and Alexandre Gramfort. Data augmentation for learning predictive models on EEG: a systematic comparison. *Journal of Neural Engineering*, 19(6):066020, 2022.

[38] Li Sun, Junxiang Chen, Yanwu Xu, Mingming Gong, Ke Yu, and Kayhan Batmanghelich. Hierarchical Amortized GAN for 3D High Res Medical Image Synthesis. *IEEE JBHI*, 2022.

[39] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[40] Mathias Perslev, Sune Darkner, Lykke Kempfner, Miki Nikolic, Poul Jorgen Jennum, and Christian Igel. U-sleep: resilient high-frequency sleep staging. *NPJ dig medicine*, 4(1):1–12, 2021.

[41] Bruno Aristimunha, Alexandre Janoni Bayerlein, M. Jorge Cardoso, Walter Hugo Lopez Pinaya, and Raphael Yokoingawa De Camargo. Sleep-Energy: An Energy Optimization Method to Sleep Stage Scoring. *IEEE Access*, 11:34595–34602, 2023.

[42] Yaniv Benny, Tomer Galanti, Sagie Benaim, and Lior Wolf. Evaluation metrics for conditional image generation. *Inter J. of Computer Vision*, 129:1712–1731, 2021.

[43] Chenlu Gao and Michael K Scullin. Longitudinal trajectories of spectral power during sleep in middle-aged and older adults. *Aging Brain*, 3:100058, 2023.

[44] Joseph Turian and Max Henry. I'm sorry for your loss: Spectrally-based audio distances are bad at pitch, 2020.

[45] Guo-Qiang Zhang, Licong Cui, Remo Mueller, Shiqiang Tao, Matthew Kim, Michael Rueschman, Sara Mariani, Daniel Mobley, and Susan Redline. The National Sleep Research Resource: towards a sleep data commons. *JAMIA*, 25(10):1351–1358, 2018.

[46] Vamsi Kumar, Likith Reddy, Shivam Kumar Sharma, Kamalaker Dadi, Chiranjeevi Yarra, Raju S. Bapi, and Srijithesh Rajendran. mulEEG: A multi-view representation learning on EEG signals. In *MICCAI*, pages 398–407. Springer, 2022.

[47] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7:267, 2013.

[48] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[49] Adam Paszke, Sam Gross, and Massa *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, pages 8024–8035. Curran Associates, Inc., 2019.

[50] M. Jorge Cardoso *et al.* MONAI: An open-source framework for deep learning in healthcare. *arXiv*, 2022.

A Further experimental settings

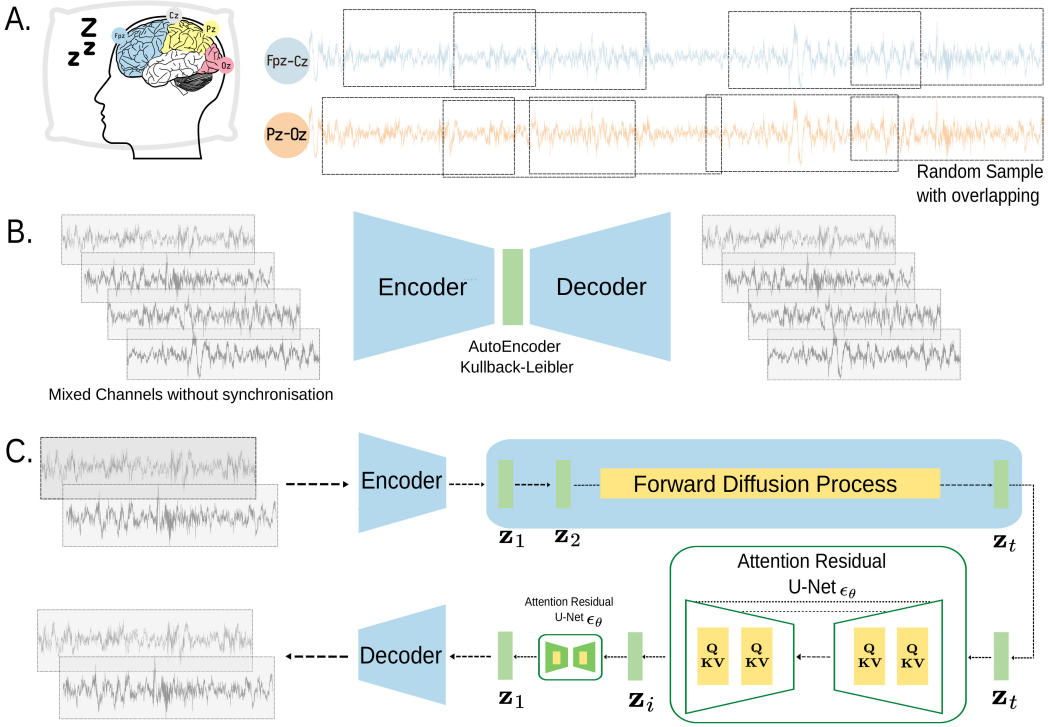

Figure 2: **Overview of our method. A.** Data sampling from the Polysomnographic recording. For each record, we sampled 30-second windows from each electrode. Every 30-second window was then used as a separate entry and combined with other subjects to create the dataset. For illustrative purposes, this figure depicts the *Sleep EDFx* electrodes, but the *SHHS* electrodes are in a similar location. **B.** Training of the AutoEncoder with Kullback-Leibler regularization and/or Spectral component to generate a compressed feature space $\mathbf{z}$ (green) **C.** The compressed feature space $\mathbf{z}$ is then used to train the Latent Diffusion model (LDM) with Attention Residual U-Net ϵ_θ (yellow). Once the LDM has been trained, the AutoEncoder decoder transforms back from the compressed representation ($\mathbf{z}$) to the EEG signals.

A.1 EEG pre-processing and epoching

In our work, we employed two extensive Sleep Stage datasets to train our deep generative model: the Physionet Sleep-EDFx dataset (*Sleep EDFx*)[1] [30] and Sleep Heart Health Study (*SHHS*)[2] [31]. These publicly available datasets offer high-quality data from large cohorts of individuals and have been utilized in various machine learning applications, including automatic sleep stage classification and analysis of sleep-related physiological signals [40, 37, 41].

In the *Sleep EDFx* dataset, we selected the Sleep Cassette (SC) subset, which comprises 153 complete sleep records from a healthy cohort of 78 subjects. We utilized the two EEG channels (Fpz-Cz and Pz-Cz), recorded with a sampling rate of 100 Hz. The *SHHS* dataset encompasses data from over 6,441 individuals who participated in a large-scale epidemiological study on sleep and cardiovascular health [45, 31]. To ensure consistency in our analyses, we resampled the *SHHS* dataset to 100 Hz. Additionally, following the methodology employed in previous works [46], we created a subsample consisting of 326 subjects, specifically selecting those with more regular sleep cycles, the $SHHS_h$ with h healthy. We utilized the two EEG channels available, C3-A2 and C4-A1.

We also applied a low-pass filter at 18 Hz in all datasets using a 5th-order Butterworth filter from the MNE-PYTHON library [47]. Higher frequencies in brain time series are typically less connected to

[1] https://physionet.org/content/sleep-edfx/1.0.0/
[2] https://sleepdata.org/datasets/shhs

sleep stage-related activity. This upper-frequency limit was carefully chosen based on an extensive investigation of Sleep Spindles, which are critical patterns in sleep stages [36]. Even though we are more focused on analyzing the lower frequencies, according to the MNE library, it is important to consider a delay compensated by the limiting edge for the band filter. Furthermore, we performed min-max standardization at the channel level to normalize all the feature spaces before inputting them into our models, leveraging the SCIKIT-LEARN library [48].

A.2 Models details

A.2.1 AutoEncoders with KL

AutoEncoders with Kullback-Leibler (KL) regularization are multi-layer networks trained to reconstruct input values ($\widehat{\mathbf{x}}$) at the output layer, with a bottleneck layer containing only a few neurons to prevent a direct copy of the input, as illustrated in Figure 2**B**. Our AE-KL architecture incorporated Residual blocks and an attention mechanism [17], with the bottleneck represented by the distribution variables μ and σ. We used two residual blocks without the attention mechanisms and two Group norm layers with the same number of dimensions as input. The trained encoder mapped the 30s windows to latent representations (denoted as $\mathbf{z}$) of size $n \times 768$, where n is the size of the mini-batch.

A.3 Forward and Reverse for the LDM

Forward Diffusion Process. Formally, as illustrated in Figure 2**C**, the forward diffusion process is described as follows: Given a feature space $\mathbf{z}_0$, the DDPM adds small increments of Gaussian noise over T steps, resulting in a sequence of noisy samples $\{\mathbf{z}_0, \ldots, \mathbf{z}_T\}$. A predefined variance schedule determines the step sizes:

$$q(\mathbf{z}_t|\mathbf{z}_{(t-1)}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1-\beta_t} \cdot \mathbf{z}_{(t-1)}, \beta_t \cdot \mathbf{I}), \tag{2}$$

$$q(\mathbf{z}_{1:T}|\mathbf{z}_0) := \prod_{t=1}^{T} q(\mathbf{z}_t|\mathbf{z}_{(t-1)}), \tag{3}$$

where, β_t is a hyperparameter defined in $\{\beta_t\}_{t=1}^{T} \in (0, 1)$, I is the identity matrix, and $\mathcal{N}(\mathbf{z}, \mu, \sigma)$ consists of a normal distribution with mean μ and covariance σ. It is important to note that the forward diffusion process does not require any trainable parameters, which makes it computationally efficient and easy to implement, resulting in faster methods [13].

Reverse Process. To reconstruct the original data point $\mathbf{z}_0$ from the Gaussian noise input $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$, the DDPM utilizes a model p_θ to approximate the conditional probabilities of the reverse diffusion process, as described below:

$$p_\theta(\mathbf{z}_{0:(T-1)}|\mathbf{z}_T) := \prod_{t=1}^{T-1} p_\theta(\mathbf{z}_{(t-1)}|\mathbf{z}_t), \tag{4}$$

$$p_\theta(\mathbf{z}_{(t-1)}|\mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{(t-1)}; \mu_\theta(\mathbf{z}, t), \Sigma_\theta(\mathbf{x}_t, t)), \tag{5}$$

In the above equations, p_θ aims to approximate the conditional probabilities of the reverse diffusion process, which reconstructs the original data point $\mathbf{z}_0$ from the final noisy sample $\mathbf{z}_T$. This is achieved by maximizing the likelihood of the reverse process given the forward process. Since the data likelihood is intractable, the model is trained using a variational lower bound on the log-likelihood.

A.4 Training setting

The training of all deep learning models was carried out using PYTORCH [49], MONAI CORE [50], and MONAI GENERATIVE [34] libraries. We employed a hold-out method to train and evaluate our generative model. We allocated 60% of the data for training, 20% for validation, and 20% for testing purposes. To process the data, we treated each EEG channel as an individual time series and randomly sampled 30-second windows for each subject. The sampled windows were then used to

build a mini-batch of 1024 windows using the MONAI library [50], as illustrated in Figure 2 **A**. For each dataset, we train and evaluate the set of electrodes together without distinction between electrodes in the sampling process.

As these random windows samples may introduce discontinuity, we applied a board padding technique by adding 72 (36 left and 36 right) points with constant values of 0 and transforming the size of the time series into a multiple of the power of 2 with windows size of 3072. This step helps mitigate potential numerical instability issues.

B Further experimental results

B.1 Samples signals

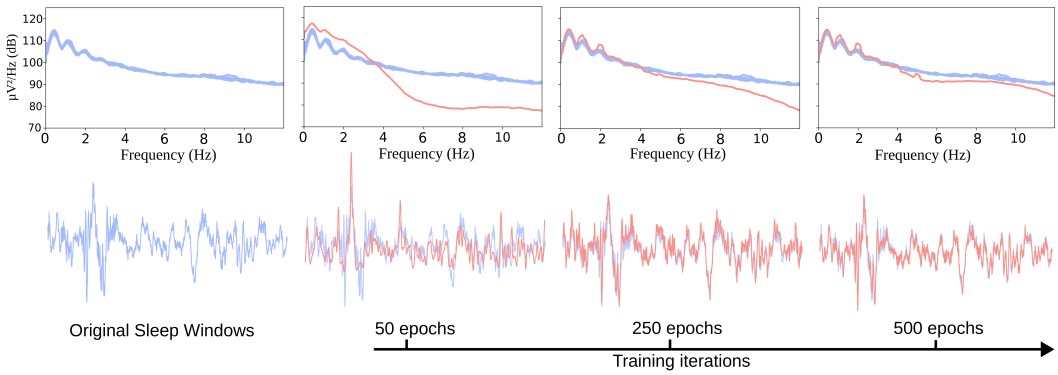

Figure 3: Evaluation of the reconstruction quality over the epochs for the AE-KL trained without the Spectral Loss. While the top row shows the Power Spectrum density of the original signal (blue) and the AE-KL reconstruction (red) after 50, 250, and 500 epochs, the bottom row illustrates the temporal oscillations of the signal. The signal reconstruction at the first epochs (epoch 50) is smoothed, and as the training progresses with more epochs, we can see that the model learns the high-frequency components of the original sleep signal.

The next three tables (Table 2 - 4) contain a breakdown of the results reported in Table 1 broken down by the stages of sleep.

Table 2: Quantitative evaluation of synthetic sleep stages on delta (δ) signal

Dataset	FID ↓			MS-SSIM ↑				
	LDM_{spec}	LDM	Real	LDM_{spec}	LDM	Real	Rec	Rec_{spec}
Sleep EDFx	5.422	**3.812**	2.912	0.453	**0.622**	0.625	**0.996**	0.964
SHHS_h	6.422	**2.021**	1.4682	**0.534**	0.461	0.503	**0.995**	0.943

Table 3: Quantitative evaluation of synthetic sleep stages on theta (θ) signal

Dataset	FID ↓			MS-SSIM ↑				
	LDM_{spec}	LDM	Real	LDM_{spec}	LDM	Real	Rec	Rec_{spec}
Sleep EDFx	**11.005**	101.162	0.6968	**0.901**	0.558	0.846	**0.994**	0.939
SHHS_h	**68.595**	282.230	0.2233	**0.949**	0.416	0.694	**0.996**	0.863

Table 4: Quantitative evaluation of synthetic sleep stages on alpha (α) signal

Dataset	FID ↓			MS-SSIM ↑				
	LDM_{spec}	LDM	Real	LDM_{spec}	LDM	Real	Rec	Rec_{spec}
Sleep EDFx	**8.185**	51.106	0.5147	**0.996**	0.547	0.932	**0.992**	0.962
SHHS_h	**43.745**	595.601	0.2697	**0.996**	0.418	0.810	**0.977**	0.891