EPiC Series in Technology

Volume 3, 2026, Pages 57–64

Proceedings of AI for Accelerated Research Symposium

Revisiting SUDEP Risk Prediction via Data Augmentation

Meiyu Li[1,2], Juliana Laze[3], Daniel Friedman[3,4], Orrin Devinsky[3,4,5,6], Zhe Sage Chen[5,6,7,*]

[1]Center for Data Science, New York University, New York, New York
[2]College of Information, University of Maryland, College Park, Maryland
[3]Comprehensive Epilepsy Center, New York University Langone Health, New York, New York
[4]Department of Neurology, New York University Grossman School of Medicine, New York, New York
[5]Neuroscience Institute, New York University Grossman School of Medicine, New York, New York
[6]Department of Psychiatry, New York University Grossman School of Medicine, New York, New York
[7]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, New York
* Corresponding author: zhe.chen@nyulangone.org

Abstract

Sudden unexpected death in epilepsy (SUDEP) is the leading cause of epilepsy-related mortality. Low-cost and noninvasive interictal biomarkers of SUDEP risk can help clinicians identify high-risk patients and initiate preventive actions. However, the small sample size in SUDEP patients remains a bottleneck for discriminatory analysis or biomarker discovery. Machine-driven data augmentation (DA) techniques can potentially alleviate the sample insufficiency or imbalance problem using synthetic data. Here we revisit an old SUDEP risk prediction problem from a new DA and generative artificial intelligence (AI) perspective, using a multicenter cohort study consisting of multichannel interictal electroencephalography (EEG) and electrocardiography (ECG) data from SUDEP patients and age-matched living epilepsy patient controls. Our results show that DA strategies can not only significantly improve the cross-validated prediction accuracy but also generalize well in newly collected held-out data samples.

1 Introduction

SUDEP is the leading cause of epilepsy-related mortality but is infrequent (~1 per 1000 patients per year) (Tomson et al., 2008; Surges et al., 2009; Devinsky, 2011). Low cost and EEG-based interictal biomarkers can potentially identify high-risk patients to enable preventive strategies (Devinsky et al., 2016). However, small samples of SUDEP patients limit development of data-driven machine learning strategies to discover statistically and clinically significant biomarkers (Ryvlin et al., 2019). Machine learning and data augmentation (DA) techniques can help address this challenge (Chen et al., 2021; Rommel et al., 2022; Chen et al., 2022b; Chen et al., 2025). Sequence modeling and generative AI

approaches, such as recurrent neural networks (RNNs), generative adversarial network (GAN), generative pre-trained transformer (GPT), and probabilistic diffusion models (DMs), have been developed to model physiological time series such as EEG (Yoon et al., 2019; Geng et al., 2021; Bird et al., 2021; Chen et al., 2024).

We used generative AI and sequence modeling approaches to develop an EEG-based DA strategy using a bi-directional long short-term memory (Bi-LSTM) model (Schuster and Paliwal, 1997). Using a previously reported multi-center SUDEP cohort (Chen et al., 2022a), we designed a principled approach to generate synthetic multi-channel EEG signals for each existing subject, we applied Bi-LSTM to produce augmented multi-channel EEG signals with equal duration per subject per condition. To identify interictal biomarkers, we first computed interictal EEG and ECG candidate features, including EEG relative power ratio and complexity features computed from spatially grouped nine EEG channels and heart rate variability (HRV) derived from one ECG lead. We trained machine learning classifiers with augmented EEG data and reported cross-validated performance. After comparing other standard DA methods, we finally validated all models using an independent held-out dataset. Finally, we discuss our findings and potential extensions of our work.

2 Methods

2.1 SUDEP cohort

This multicenter, retrospective, case-control study identified SUDEP cases among patients admitted to eight epilepsy monitoring units (EMUs) of the MS-BioS Study Group (including the Royal Melbourne Hospital, Austin Hospital, St. Vincent's Hospital, Melbourne, Australia; NYU Langone Health, NY Presbyterian Hospital/Columbia University, New York; University of Cincinnati, Cincinnati; Yale New Haven Hospital, New Haven; and Johns Hopkins Medical Center, Baltimore). Patients who underwent video EEG monitoring (VEM) with >21-scalp electrodes using the 10-20 System and lead II of a standard 12-lead ECG.

Each center identified patients aged 6 months to 65 years with ≥1 electroclinical seizure recorded over a 2-11-year consecutive period. All patients were followed for ≥5 years. Epilepsy-related deaths were reviewed with available records, medical examiner/coronial and autopsy findings to determine cause of death. Definite and probable SUDEP cases were included based on current criteria. For each SUDEP case, two living epilepsy controls were matched according to admission age (±4 years), sex, EMU admission year (±1 year) from the EMU cohort at each center. Epilepsy controls had documented contact in the medical record within six months of screening or were identified as not deceased from national death records.

In the prior study (Chen et al., 2022a), the multicenter cohort consisted of 88 participants. A subset comprising of 83 subjects, including 30 SUDEP patients (14 female [47%]; mean age [SD], 31 [8.47] years) and 58 living epilepsy controls (26 female [45%]; mean age [SD] 31 [8.5] years), had 10-min interictal sleep EEG recordings (note: in one SUDEP case and four controls, interictal EEG recordings during sleep were unavailable). Furthermore, 76 subjects (26 SUDEP and 50 controls) from this subset had concurrent interictal EEG and ECG recordings from wake and sleep periods. In the newly expanded cohort (up to October 2024), 27 more new subjects (8 SUDEP and 19 controls) with similar recording settings have been added, which were collected from five additional EMU centers from Northwestern University, Italy, Spain, Germany, and Austria. These datasets will serve as an independent held-out validation set for the prior training strategy.

2.1 Data recordings

In the SUDEP cohort, the sampling rate of EEG signals ranged between 256 and 512 Hz. For each subject, 10-min interictal segments from non-rapid-eye-movement (NREM) sleep and 10-min segment from wakefulness were identified during video EEG monitoring (VEM). Sleep segments were chosen at random that preceded at least 1 hour before or after a non-convulsive seizure or 6 hours before or after a convulsive seizure. During wakefulness, EEG signals were collected when subjects were at free of muscle and movement artefact. These artefact-free segments were chosen by a well-trained epileptologist who had no knowledge of any hypothesis. We excluded the immediate post-ictal period, which was defined as ≥6 hours following tonic-clonic seizures and ≥1 hour for all other seizure types. Additionally, 10-min interictal segments of stable concurrent ECG recordings were selected from both NREM sleep and wakefulness for each subject.

2.2 EEG and ECG-HRV features

EEG signals were preprocessed with a notch filter around 60 Hz and a band pass filter from 1-100 Hz. The MNE Python library (https://mne.tools/) was used for preprocessing. For each EEG group signal, we performed band-pass filtering (1-100 Hz) and then computed the relative power ratio at six frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-15 Hz), beta (15-30 Hz), low gamma (30-50 Hz), high gamma (50-100 Hz). The spectral power was calculated using fast Fourier transform (FFT) with a multi-taper estimator on the entire 10-min recordings. To calibrate the EEG measurement variability between subjects, we computed the relative power percentage and sleep-to-wake power ratio (Chen et al., 2022a). Furthermore, we spatially clustered scalp EEG electrode channels into nine groups (G1-G9) to reduce the feature dimensionality: G1 ('FP1', 'F7', 'F9'), G2 ('FP2', 'F8', 'F10'), G3 ('FZ', 'F3', 'F4'), G4 ('T7', 'T9', 'A1'), G5 ('T8', 'T10', 'A2'), G6 ('CZ', ' C3', 'C4'), G7 ('P7', 'P9', 'O1'), G8 ('P8', 'P10', 'O2'), G9 ('PZ', 'P3', 'P4 '). In total, we computed 6×9=54 (frequency×group) power ratio features per subject.

In addition, we computed the following nonlinear or complexity features: Lempel-Ziv complexity, detrended fluctuation analysis (DFA), Hjorth activity and complexity (https://github.com/JingweiToo/EEG-Feature-Extraction-Toolbox), and sample entropy. We also computed the aperiodic (1/f-like) exponent of the EEG spectrum (Donoghue et al., 2020) using the "Fitting Oscillations and One-Over-f" (FOOOF) toolbox (https://fooof-tools.github.io/fooof/). Together, we considered the sleep-to-wake ratios of 6 EEG nonlinearity features.

To characterized heart rate variability (HRV) of ECG signals, we used a software package (https://github.com/Aura-healthcare/hrv-analysis) to compute a set of temporal and spectral features as well as nonlinear features (a total of 24 HRV candidate features per subject). For the calibration purpose, we again computed each HRV feature's sleep-to-wake ratio so that the actual feature was dimensionless.

As described in our prior publication (Chen et al., 2022a), we used a nested feature selection strategy to identify the most relevant EEG and ECG feature for discriminatory analysis. Among three feature blocks, we conducted feature selection on a single feature block basis. Upon finding the most discriminative features from each block, we then combined all features from three blocks and ran the cross-validated classification analysis, from which we further determined the relative importance of these features.

2.3 Data augmentation

DA may occur at the EEG feature level or raw signal level. Our proposed approach was operated at the raw signal level. Specifically, we used a bidirectional LSTM (Bi-LSTM) for augmenting EEG time series for each electrode channel per subject. For each electrode channel, we first divided the EEG time series sequence (with length 100,000 to 300,000) into 1,000 pieces (**Fig. 1A**). For the sequence in each piece, we used Bi-LSTM to generate synthetic data and concatenated them together. After that we

treated it a synthetic duplicate for that specific channel and specific subject and assigned the same label (1 for SUDEP and 0 for control) as the original subject. Bi-LSTM consists of two LSTM layers, one for processing the input sequence in the forward direction and the other in the backward direction. Each LSTM layer consists of a set of building blocks of "LSTM cells" that use a 'tanh' activation function for the cell state and uses a 'sigmoid' activation function for the node output. Both LSTM layers return a probability vector as output and the final output is the combination of these two probabilities (**Fig. 1B**): $p_t = p_t^f + p_t^b$, where p_t^f and p_t^b denote the probability vector from the forward and backward LSTM layers, respectively. The dimension of input in Bi-LSTM is [20, 5, 1] where 20 is the batch size, 5 is the loopback period (window size), 1 is the number of electrodes. We set the number of layers as 1, hidden size as 50. A linear layer was added after Bi-LSTM to the output unified dimension.

To generate realistic yet sufficiently different synthetic data, each piece of the EEG signals was fed into Bi-LSTM as the input, and the Bi-LSTM generated synthetic EEG signals in the output. At first, we split the first 60% time series as the training set and the remaining 40% as the testing set. We further transformed the time series in each set into a tensor dataset with a window size of 5. That is, the transformed dataset used [EEG_t, EEG_{t+4}] as the features and [EEG_{t+1}, EEG_{t+5}] as the labels, where EEG_t denotes the EEG value at time *t*. Next, we trained Bi-LSTM using transformed training data for 5 epochs. For each epoch, mini batches of training data were iterated with a batch size of 20. In each mini-batch, the model parameters were updated using the Adam optimizer based on the gradient of the mean-squared error (MSE) between the predicted and actual target values. A validation step was also performed at each epoch. We chose RMSE to compute the training and validation losses. The predicted EEG signal sequence with the minimal RMSE in the validation loss was selected.

As method comparison, we also implemented a few other standard DA approaches for EEG signals (He et al., 2021). At the feature level, we considered two representative approaches, one based on adding Gaussian background noise, and the other based on SMOTE (Synthetic Minority Oversampling Technique) (Chawla et al., 2002). At the raw signal level, one approach was based on adding Gaussian noise to original EEG recordings for each electrode (Rommel et al., 2022). We also adopted two established EEG data augmentation approaches for the classification problem: Smooth Time Masking (STM) and Amplitude Scale (AS) (Mohsenvand et al., 2020). The STM method was designed to replace a part of EEG signals by 0s. In our experiment, we masked randomly 20% of EEG sequence to be 0. For the AS method, we scaled our EEG signal by 0.5 to 2 using a similar strategy described in the paper (Mohesenyand et al., 2020). All the DA methods have the same augmented sample size.

To create sample-matched ECG-HRV features associated with the augmented EEG data, we used a bootstrapping method to resample the data with replacement.

2.4 Machine learning

Here we tested two standard machine learning classifiers, including the regularized L_1 logistic regression (LR) classifier and linear support vector machine (SVM). In all classifiers, the binary label 0/1 represents the non-SUDEP/SUDEP case. The machine learning classifiers were implemented from the scikit-learn library (https://scikit-learn.org/). We computed the AUC (area under the receiver operating characteristic curve) for each machine learning classifier. Median AUC and IQR (25% and 75% percentiles) were computed using 5-fold cross validation with 1000 random repeats. Given the synthetic EEG data, we used the augmented (i.e., original plus synthetic) data only in training and kept the testing setup unchanged for fair comparison. Upon cross-validating DA in the training set, we retrained the classifiers with all available training samples (n=99) and further validated the model with the independent held-out dataset (n=27).

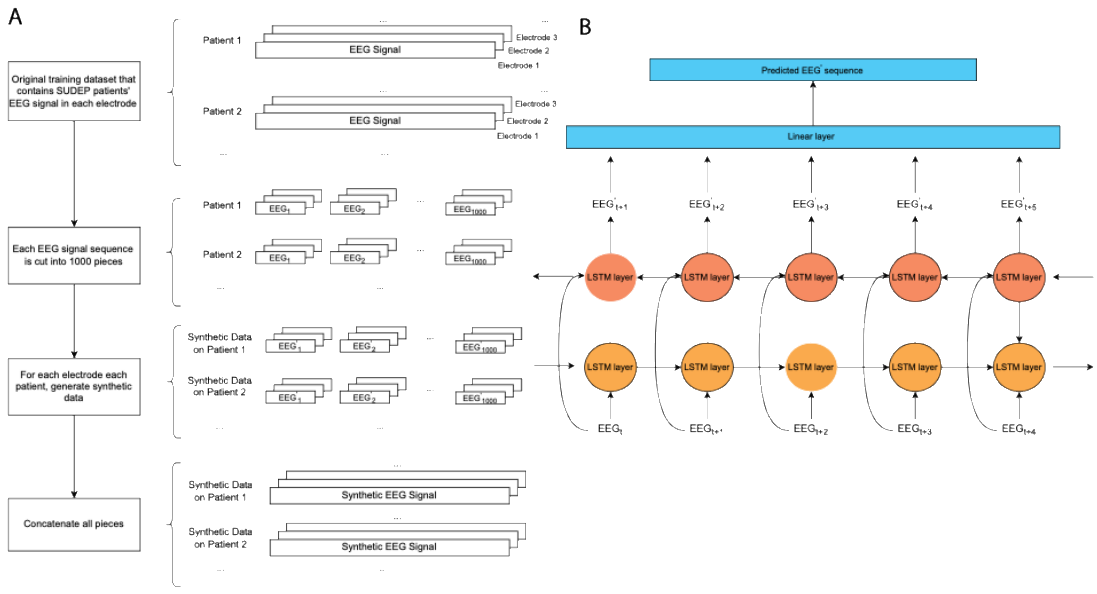

Figure 1. (A) Flowchart of EEG data augmentation. (**B**) Schematic of Bi-LSTM for generating synthetic EEG segments.

3 Results

To conduct EEG data augmentation for each EEG channel, we divided 10-min EEG into 0.6-s long segments (1200 samples) and applied Bi-LSTM to generate synthetic EEG signals; we then concatenated all segments together and repeated the procedure for all 9 spatially grouped channels (**Fig. 1A**). For each subject, we generated synthetic EEG time series (and corresponding features). To balance the class sample size, we used only the synthetic data from SUDEP patients for subsequent classification analysis. It should be emphasized that inclusion of the synthetic data from controls did not improve the performance (results not shown). Using the proposed Bi-LSTM DA strategy, we improved the classification accuracy by comparing the same classifiers (**Table 1**), increasing the AUC performance by ~5% from the previously published result (Chen et al., 2022). The improvement was small but significant (P<0.05, Wilcoxon rank-sum test), considering the result was evaluated on the same amount of real data. Additionally, we found that the simplest LR classifier yielded slightly yet consistently better performance than the SVM and random forest classifiers.

All DA strategies tested thus far showed somewhat improvement in cross-validated AUC, but the Bi-LSTM-based DA strategy outperformed other EEG-based DA approaches (**Table 2**). During independent test data validation, we found that the AUC result (0.80) was remarkably on a par with the cross-validated AUC statistic (mean 0.81), suggesting excellent out-of-sample generalization based on our selected features and classifiers. In addition to the AUC, we also obtained Sensitivity (Recall) of 0.714, Specificity of 0.765, Accuracy of 0.750, Precision of 0.556, and F1 score of 0.625 for the held-out data. Finally, we also tested a hybrid DA strategy by combining our DA approach with another independent DA method. However, the hybrid strategy did not show any further AUC improvement, implying an information bottleneck among the existing data.

Feature set	Classifier	#SUDEP+ #Controls (w/o DA)	AUC (w/o DA)	#SUDEP+ #Controls (DA: Bi-LSTM)	AUC (DA: Bi-LSTM)
EEG alone	LR	n = 76 (26 + 50)	0.75 [0.73, 0.78]	n =99 (49 + 50)	0.77 [0.74, 0.79]
	SVM		0.74 [0.68, 0.78]		0.77 [0.70, 0.79]
EEG+ECG	LR	n = 70 (24 + 46)	0.77 [0.73, 0.80]	n = 93 (47 + 46)	**0.81 [0.78, 0.84]**
	SVM		0.74 [0.70, 0.78]		0.78 [0.75, 0.82]

Table 1. Comparison of AUC results (median [IQR]) based on our proposed Bi-LSTM data augmentation (DA) method.

Method	5-fold cross-validated AUC (median [IQR]), n=70	AUC on independent held-out data (n=27)
w/o DA	0.77 [0.73, 0.80]	0.77
DA: Bi-LSTM	**0.81 [0.78, 0.84]**	**0.80**
DA: SMOTE	0.78 [0.74, 0.81]	0.77
DA: Gaussian noise addition	0.76 [0.72, 0.77]	0.75
DA: Smooth Time Masking	0.78 [0.72, 0.81]	0.78
DA: Amplitude Scaling	0.79 [0.72, 0.84]	0.78

Table 2. Result comparisons between various data augmentation (DA) methods based on the LR classifier.

4 Discussion and conclusion

In this paper, we have aimed to identifying robust interictal biomarkers for SUDEP based on an expanded retrospective multicenter cohort. Extending our prior effort (Chen et al., 2022a), here we focus on synthetic data generation, which has been recently developed rapidly in the data-centric AI field to change the landscape of biomedical research. In SUDEP research, limited data samples are severely imbalanced, creating a bottleneck for machine learning-based predictive analytics. The use of synthetic yet realistic EEG data via AI-empowered DA techniques, such as GANs, generative DMs, and generative EEG transformer, may help address these concerns (Lashgari et al., 2020; Rasheed et al., 2021; Ali et al., 2024). Currently, each spatially grouped EEG channel was treated independently. This is an oversimplified assumption as multi-channel EEG channels are also spatially coupled. Future DA strategies for capturing spatiotemporal features of multi-channel EEG may yield more accurate characterizations in synthetic data (Vetter et al., 2024). Potential considerations are probabilistic DMs (e.g., Torma and Szegletes, 2025; Chetkin et al., 2024), GANs (e.g., Habashi et al., 2023), and other cutting-edge methods such as EEGTrans (Lim and Kuo, 2025), but their intrinsic computational cost for long multichannel time series need to be mitigated.

In our current investigations, we noticed that while GPT is highly effective for text generation as it is trained on large corpus of text, it is not well-suited for signal generation tasks such as EEG. Similarly, GANs requires large datasets to balance generator and discriminator learning. With small datasets, GANs tend to fail or collapse. Additionally, synthetic data generation via generative DMs are more computationally expensive for long EEG durations (especially with higher temporal resolution). In comparison, our proposed Bi-LSTM method can be trained with a smaller dataset, and its quality generally improves with sample size, offering a good computational trade-off in quality and efficiency.

Additionally, we need to develop quantitative methods to assess the quality of synthetic EEG data (e.g., waveforms, temporal/spectral/spatial features, and functional connectivity), although there is no consensus on the right metric of synthetic EEG signals for the downstream classification task (Zhang

and Chen, 2025). Pre-trained EEG foundation models (see Chen et al., 2025 for a review) with strategic fine-tuning may improve the classification performance. We plan to pursue these in future investigations and validate the biomarkers using a much larger SUDEP cohort.

In conclusion, our results suggest that the DA strategy may provide a promising research direction for generating augmented EEG samples. Our results in interictal EEG-based SUDEP risk assessment are significant in terms of potential translational impact, and the emerging AI-driven DA methods can benefit a wide range of research in SUDEP, epilepsy, and neurological disorders.

Acknowledgments

We thank Aaron Hsieh and Ran Wang for sharing the code for data preprocessing and feature computation. We also thank the large MS-BioS group for the contribution of SUDEP cohort. This study was funded by grants from the US National Institutes of Health (R01-NS123928, R01-NS123928-02S1) and Multidisciplinary University Research Initiatives (MURI), Centers for Disease Control and Prevention (CDC), Finding a Cure for Epilepsy and Seizures (FACES), and Oracle for Research Project Award.

References

1. Abayomi-Alli, O.O., Damaševičius, R., Maskeliūnas R, Abayomi-Alli, A. BiLSTM with data augmentation using interpolation methods to improve early detection of Parkinson disease. *Proc. 15th Conf. Computer Science and Information Systems*, Sofia, Bulgaria, 371-380 (2020).
2. Ali O, Saif-ur-Rehman M, Metzler M, et al. GET: a generative EEG transformer for continuous context-based neural signals. https://arxiv.org/abs/2406.03115 (2024).
3. Bird JJ, Pritchard M, Fratini A, Ekárt, A, Faria, DR. Synthetic biological signals machine-generated by GPT-2 improve the classification of EEG and EMG through data augmentation. *IEEE Robotics and Automation Letters*, 6, 3498-3504 (2021).
4. Carrle FP, Hollenbenders Y, Reichenbach A. Generation of synthetic EEG data for training algorithms supporting the diagnosis of major depression disorder. *Front. Neurosci.* 17, 1219133 (2023).
5. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16, 321-357 (2002).
6. Chen, R.J., Lu, M.Y., Chen, T.Y. et al. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed. Eng* 5, 493-497 (2021).
7. Chen ZS, Hsieh A, Sun G, et al. Interictal EEG and ECG for SUDEP risk assessment: a retrospective multicenter cohort study. *Front. Neurology*, 13, 858333 (2022a).
8. Chen ZS, Kulkarni P, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y. Modern views of machine learning for future precision psychiatry. *Patterns*, 3, 100602 (2022b).
9. Chen ZS. Emerging brain-to-content technologies from generative AI and deep representation learning. *IEEE Signal Proc. Mag.*, 41, 94-104 (2024).
10. Chen ZS, Lu B-L, Wu W. Representation learning and foundation models for electroencephalography analyses: current trends, fundamental insights and future directions. *IEEE Signal Proc. Mag.* 42, 36-57 (2025).
11. Chetkin EI, Kozyrsky BL, and S. L. Shishkin SL. Unconditional EEG synthesis based on diffusion models for sound generation. Proc. 2024 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 416-420 (2024).
12. Devinsky O. Sudden, unexpected death in epilepsy. *N. Engl. J. Med.* 365, 1801-1811 (2011).
13. Devinsky O, Hesdorffer DC, Thurman DJ, Lhatoo S, Richerson G. Sudden unexpected death in epilepsy: epidemiology, mechanisms, and prevention. *Lancet Neurol.* 15,1075-1088 (2016).

14. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, & Voytek B. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.,* 23, 1655-1665 (2020).
15. Geng D, Alkhachroum A, Melo Bicchi M, Cajigas I, Chen ZS. Deep learning for robust detection of interictal epileptiform discharges. *J. Neural Eng.,* 18, 056015 (2021).
16. Habashi, A.G., Azab, A.M., Eldawlatly, S. et al. Generative adversarial networks in EEG analysis: an overview. *J NeuroEngineering Rehabil* 20, 40 (2023).
17. 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).
18. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans. Knowledge and Data Eng.* 21, 1263-1284 (2009).
19. Lhatoo SD, Faulkner HJ, Dembny K, Trippick K, Johnson C, Bird JM. An electroclinical case-control study of sudden unexpected death in epilepsy. *Annals of Neurology.* 68, 787-796 (2010).
20. Lashgari, E., Liang, D., Maoz U. Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods*, 346, 108885 (2020).
21. Lim, J-H., Kuo, P-C. EEGTrans: Transformer-driven generative models for EEG synthesis. *Int. Conf. Learned Representations (ICLR),* (2025).
22. Mohsenvand M, Izadi MR, Maes P. Contrastive representation learning for electroencephalogram classification. *Proc. Machine Learning Res*. 136, 238-253 (2020).
23. Rasheed K, Qadir J, O'Brien TJ, Kuhlmann L, Razi A. A Generative model to synthesize EEG data for epileptic seizure prediction. *IEEE Trans Neural Syst Rehab. Eng*. 29, 2322-2332 (2021).
24. Rommel C, Paillard J, Moreau T, Cramfort A. Data augmentation for learning predictive models on EEG: a systematic comparison. *J. Neural Eng.* 19 066020 (2022).
25. Ryvlin P, Rheims S, Lhatoo SD. Risks and predictive biomarkers of SUDEP. *Curr. Opin. Neurol.* 32, 205-212 (2019).
26. Schuster M., Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.*, 45, 2673-2681 (1997).
27. Surges R, Thijs RD, Tan HL, Sander JW. Sudden unexpected death in epilepsy: risk factors and potential pathomechanisms. *Nat. Rev. Neurol.* 5, 492-504 (2009).
28. Tomson T, Nashef L, Ryvlin P. Sudden unexpected death in epilepsy: current knowledge and future directions. *Lancet Neurol.* 7, 1021-1031 (2008).
29. Torma S, Szegletes L. Generative modeling and augmentation of EEG signals using improved diffusion probabilistic models. *J. Neural Eng*. 22, 016001 (2025).
30. Vetter J, Macke JH, Gao R. Generating realistic neurophysiological time series with denoising diffusion probabilistic models. *Patterns*, 5, 101047 (2024).
31. Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. *Advances in Neural Information Processing Systems (NeuroIPS),* pp. 5508-5518 (2019).
32. Zhang Y, Chen ZS. Harnessing electroencephalography connectomes for cognitive and clinical neuroscience. *Nat. Biomed. Eng.*, 9, 1186-1201 (2025).