End-to-End Diffusion Modeling for Clinical EEG Abnormality Detection with Spatial Filtering and Attention

Jae-Pil Lee [1] **Jin-Young Chung** [1] **Wiha Choi** [2] **Jaeho Jang** [2] **Hyang-Sook Hoe** [2 3] **Hyun-Chul Kim** [1]

Abstract

Automated interpretation of clinical electroencephalography (EEG) is challenging due to signal heterogeneity, noise contamination, and inter-subject variability. We propose DiffSA-EEG, a diffusion-based EEG classification framework that integrates learnable spatial filtering, stacked denoising autoencoders (SDA), and convolutional block attention modules (CBAM) within an end-to-end discriminative pipeline. Unlike prior diffusion-based EEG studies that focus on data generation or augmentation, our framework leverages denoising diffusion probabilistic models (DDPMs) directly as an auxiliary denoising objective within a discriminative classifier. We evaluate DiffSA-EEG on two large-scale clinical EEG corpora: the Temple University Hospital Abnormal EEG Corpus (TUAB) and the Temple University Epilepsy Corpus (TUEP). DiffSA-EEG achieves the best or comparable performance against established baselines—including EEGNet, Deep4Net, ChronoNet, temporal convolutional networks, and the Diff-E backbone—with the most consistent gains in accuracy and AUC-ROC. Ablation analyses reveal that optimal component combinations are dataset-dependent: spatial filtering with SDA is most effective for TUAB, while SDA with CBAM yields superior performance on TUEP. Grad-CAM-based interpretability analysis further shows spatial patterns broadly consistent with established neurophysiological biomarkers.

This paper was accepted to the Workshop on Structured Data for Health (SD4H) at ICML 2026. [1]Department of Artificial Intelligence, Kyungpook National University, Daegu, Republic of Korea [2]Korea Brain Research Institute (KBRI), Daegu, Republic of Korea [3]Department of Brain and Cognitive Sciences, DGIST, Daegu, Republic of Korea. Correspondence to: Hyun-Chul Kim <hyunchul_kim@knu.ac.kr>.

Proceedings of the Workshop on Structured Data for Health at the 43rd International Conference on Machine Learning, Seoul, South Korea. Copyright 2026 by the author(s).

1. Introduction

Electroencephalography (EEG) is a cornerstone non-invasive neuroimaging technique widely used for diagnosing neurological disorders such as epilepsy and stroke (Acharya et al., 2012; Obeid & Picone, 2016). However, clinical EEG interpretation remains heavily reliant on expert visual inspection, which is subjective and labor-intensive, motivating the development of automated deep learning approaches (Roy et al., 2019b).

Recent deep learning methods—including EEGNet (Lawhern et al., 2018), Deep4Net (Schirrmeister et al., 2017), ChronoNet (Roy et al., 2019a), and temporal convolutional networks (TCN) (Gemein et al., 2020)—have shown promising results on benchmark datasets such as TUAB and TUEP. More recently, transformer-based and state-space models have further advanced performance by capturing long-range temporal dependencies (Tegon et al., 2025). In parallel, denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020) have emerged as a powerful tool in biomedical signal processing. For example, Diff-E demonstrated the potential of conditional diffusion models for decoding imagined speech from EEG (Kim et al., 2023), while DiffMDD applied diffusion-based learning for major depressive disorder diagnosis (Wang et al., 2024). However, existing diffusion-based EEG studies have primarily focused on data generation or augmentation (Shu et al., 2025; Jiang et al., 2024), rather than leveraging diffusion as a core component for end-to-end discriminative classification on large-scale clinical datasets. These diffusion-based EEG models also differ from ours in objective: Diff-E (Kim et al., 2023) targets imagined-speech decoding, DiffMDD (Wang et al., 2024) addresses depression diagnosis, and EEG-DIF (Jiang et al., 2024) performs generative seizure forecasting. In contrast, DiffSA-EEG repurposes the denoising diffusion process itself as an auxiliary denoising objective for end-to-end abnormality classification on large-scale clinical corpora.

To address this gap, we propose **DiffSA-EEG**, a diffusion-based EEG classification framework that integrates: (1) a learnable *spatial filter* (SF) for topographical dimensionality reduction, (2) a *stacked denoising autoencoder* (SDA) for noise-invariant latent representations, and (3) *convolutional block attention modules* (CBAM) (Woo et al., 2018) for

selective feature emphasis. Unlike prior work, DiffSA-EEG embeds this denoising process within a unified encoder–decoder architecture, coupled with a conditional autoencoder (CAE) and a classification head.

Our key contributions are: (i) repurposing the denoising diffusion process as an auxiliary prior for discriminative EEG classification, rather than for generation, (ii) a systematic ablation study revealing dataset-dependent optimal configurations, and (iii) Grad-CAM-based interpretability analysis illustrating clinically plausible spatial patterns.

2. Methods

2.1. Framework Overview

DiffSA-EEG extends the Diff-E architecture (Kim et al., 2023) with a unified encoder–decoder framework comprising four core modules (Figure 1):

(1) DDPM Backbone. Given a clean EEG signal $X_0 \in \mathbb{R}^{C\times T}$, the forward diffusion process produces a noisy sample $X_t = \sqrt{\bar{\alpha}_t}X_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$ and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ is the cumulative signal retention ratio. The reverse process, parameterized by a time-conditional U-Net (Ronneberger et al., 2015), directly reconstructs X_0 from X_t, providing signal-level guidance for downstream tasks. Our model is trained to recover the clean signal X_0 directly rather than the injected noise (Eq. 1), so the diffusion backbone serves as an auxiliary denoising prior rather than a generator.

(2) Learnable Spatial Filter (SF). To reduce channel redundancy, a spatial filter computes $F = W_f U^\top X_0 \in \mathbb{R}^{M\times T}$, where $U \in \mathbb{R}^{C\times M}$ contains the dominant M spatial bases obtained by an SVD of the training data and $W_f \in \mathbb{R}^{M\times M}$ is a learnable projection matrix. A learned linear layer subsequently reconstructs the channel space after diffusion-guided refinement.

(3) Stacked Denoising Autoencoder (SDA). Following Vincent et al. (2010), Gaussian noise is injected into latent features ($\tilde{z}^{(i)} = z^{(i)} + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, $\sigma = 0.05$) through three stacked convolutional layers, encouraging noise-invariant representations shared by both the decoder and the classification head.

(4) CBAM. Channel and temporal attention mechanisms (Woo et al., 2018) are sequentially applied after residual convolution blocks to emphasize informative EEG features. Self-attention blocks (Vaswani et al., 2017) are additionally applied after each encoder stage to capture long-range temporal dependencies.

Conditional Autoencoder (CAE) corrects diffusion-induced reconstruction errors via skip connections from intermediate DDPM layers, trained with a Smooth L1 loss.

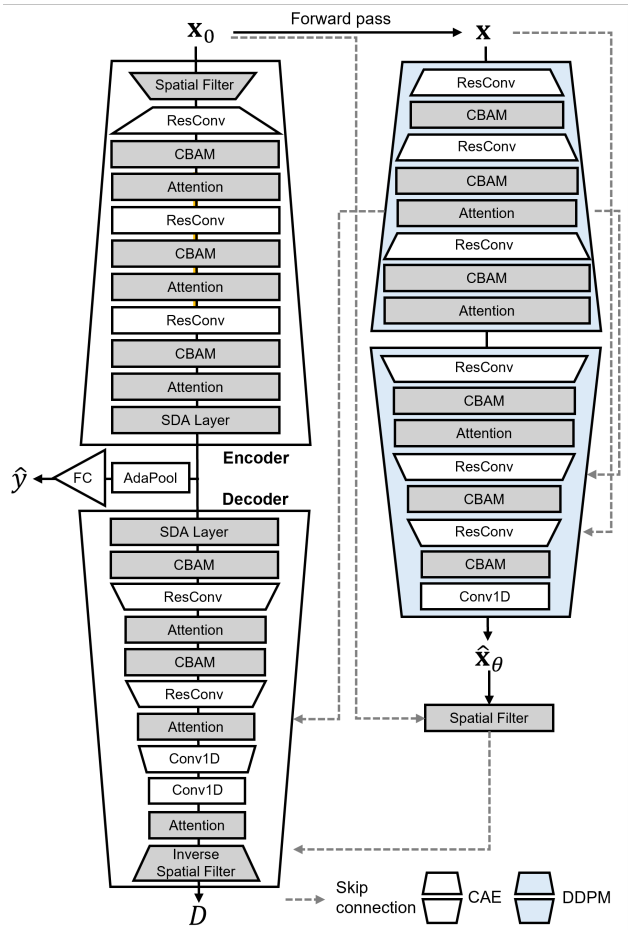

Figure 1. Overview of DiffSA-EEG. The DDPM backbone generates diffusion-refined representations, which are corrected by the CAE and classified via adaptive pooling and a fully connected head. SF, SDA, and CBAM modules are integrated throughout the encoder–decoder pathway.

Both the original input X_0 and the DDPM output $\hat{X}_0$ are injected into the final decoder layer, enabling the CAE to leverage both global structure and local refinement cues. The DDPM backbone is optimized separately (Stage 1) and then detached; its reconstruction-error signal supervises the CAE, and the discriminative branch (CAE and classification head) is trained in Stage 2. The two stages optimize:

$$\text{Stage 1:} \quad \mathcal{L}_{\text{DDPM}} = \mathbb{E}_{t,\epsilon}[\,\|X_0 - f_\theta(X_t, t)\|_1\,], \quad (1)$$

$$\text{Stage 2:} \quad \mathcal{L} = \mathcal{L}_{\text{CAE}} + \alpha\,\mathcal{L}_{\text{CLS}}, \qquad \alpha = 0.1, \quad (2)$$

where $\mathcal{L}_{\text{CAE}}$ is a Smooth L1 loss between the CAE output and the (detached) DDPM reconstruction-error map, and $\mathcal{L}_{\text{CLS}}$ is the cross-entropy classification loss.

Decoupling the two stages is deliberate. Training the diffusion backbone independently and detaching its output before the discriminative stage prevents the classification gradient from distorting the generative reconstruction tar-

Table 1. Performance comparison on TUAB and TUEP (mean ± SE). The best result in each column is shown in **bold**. Per-baseline statistical comparisons are reported in Section 3.

Model	TUAB					
	Acc (%)	Recall (%)	Spec (%)	AUC-ROC (%)	AUC-PR (%)	Bal Acc (%)
EEGNet	78.0±0.4	77.7±0.4	74.4±1.4	83.3±0.3	55.0±5.0	46.8±5.0
Deep4Net	82.5±0.6	82.0±0.5	76.0±1.1	89.5±0.4	**90.0±0.4**	**81.9±0.5**
ChronoNet	81.7±0.7	81.2±0.7	75.0±0.8	87.8±0.8	88.4±0.8	81.2±0.8
TCN	79.9±0.1	79.3±0.2	72.1±1.1	84.7±0.2	84.7±0.1	79.3±0.2
Diff-E	82.6±0.1	81.8±0.1	72.1±0.8	88.2±0.1	87.6±0.2	76.9±0.5
DiffSA-EEG	**83.9±0.3**	**83.6±0.3**	**79.4±0.8**	**90.6±0.3**	89.4±0.5	81.5±0.6

Model	TUEP					
	Acc (%)	Recall (%)	Spec (%)	AUC-ROC (%)	AUC-PR (%)	Bal Acc (%)
EEGNet	58.1±1.8	58.3±1.8	17.6±3.8	90.0±1.1	54.5±2.9	46.8±5.0
Deep4Net	85.2±1.1	85.3±1.1	71.2±2.3	96.7±0.3	94.7±0.9	81.9±0.5
ChronoNet	90.4±0.3	90.4±0.3	88.2±1.1	96.5±0.2	96.0±0.3	81.2±0.8
TCN	91.6±0.1	91.6±0.1	87.9±0.4	96.5±0.1	93.8±0.3	79.3±0.2
Diff-E	94.6±0.1	94.6±0.1	92.0±0.5	98.1±0.1	97.6±0.1	93.3±0.3
DiffSA-EEG	**94.7±0.2**	**94.7±0.2**	**93.2±0.2**	**98.3±0.1**	**97.8±0.3**	**94.0±0.2**

get, so the diffusion module provides a stable, task-agnostic denoising prior. The CAE then learns to predict the residual reconstruction error of this frozen prior, allowing the discriminative branch to exploit diffusion-refined features without back-propagating through the full sampling trajectory. This design keeps the generative and discriminative objectives from competing, which we found important for stable optimization on noisy clinical recordings.

2.2. Datasets and Preprocessing

TUAB (v3.0.1) comprises 2,993 EEG recordings (1,521 normal, 1,472 abnormal) from 2,130 patients diagnosed with conditions such as epilepsy, stroke, depression, and Alzheimer's disease. **TUEP** (v2.0.1) contains 2,298 recordings from 200 patients (513 non-epileptic, 1,785 epileptic); after exclusion criteria (duration $<$ 130s or missing channels), 317 non-epileptic and 1,275 epileptic recordings remained, with k-means SMOTE (Last et al., 2018) applied to the full TUEP set for class balancing prior to partitioning, rather than to the training split alone. Both datasets used 21 channels (international 10–20 system), sampled at 250 Hz, with 1 Hz high-pass Butterworth and 60 Hz notch filters applied. The first 60 seconds were segmented with time-reversal augmentation (Rommel et al., 2022), yielding two non-overlapping segments per recording, normalized via z-scoring.

3. Experiments and Results

All models were trained for 100 epochs across 10 independent runs with different random seeds on an NVIDIA RTX A6000 GPU. For TUAB, official train/test splits were used; for TUEP, 80/20 random splits were repeated 10 times. Statistical significance was assessed with a bootstrap-based two-sample test of the difference in means (10,000 resamples), comparing DiffSA-EEG to each baseline. All results are reported as mean ± standard error (SE).

Implementation details. The DDPM backbone uses a cosine noise schedule ($s = 0.008$) with $T = 1000$ diffusion timesteps, and the latent SDA perturbation is fixed at $\sigma = 0.05$. All models are optimized with AdamW (weight decay 10^{-4}) under a cyclic learning-rate schedule ranging from 10^{-5} to 4×10^{-3}, with a batch size of 32. As described in Section 2, the diffusion backbone is trained first and detached; the discriminative branch (CAE and classifier) is then optimized with the combined objective, in which the classification term carries weight $\alpha = 0.1$. An exponential moving average of the classifier parameters is maintained for evaluation. These settings are shared across all baselines and ablation configurations to ensure a controlled comparison.

Table 2. Top ablation configurations on TUAB and TUEP (mean ± SE). Full model includes all four modules.

Configuration	TUAB			TUEP		
	Acc (%)	AUC-ROC (%)	Bal Acc (%)	Acc (%)	AUC-ROC (%)	Bal Acc (%)
Diff-E (base)	82.6±0.1	88.2±0.1	76.9±0.5	94.6±0.1	98.1±0.1	93.3±0.3
SF + SDA	**84.6±0.1**	**90.8±0.2**	**81.6±0.6**	94.9±0.2	98.5±0.1	94.6±0.3
SDA + CBAM	82.1±0.2	87.9±0.3	79.7±0.8	**96.0±0.2**	98.4±0.2	**95.1±0.2**
SF + SDA + CBAM	84.2±0.3	90.6±0.3	82.6±0.6	95.4±0.2	**98.6±0.1**	94.8±0.3
Full (all four)	83.9±0.3	90.6±0.3	81.5±0.6	94.7±0.2	98.3±0.1	94.0±0.2

3.1. Comparison with Baselines

Table 1 presents the comprehensive comparison between DiffSA-EEG and five baseline models across six evaluation metrics. On TUAB, DiffSA-EEG attained the highest accuracy (83.9%), specificity (79.4%), and AUC-ROC (90.6%) among all models. On accuracy, recall, specificity, and AUC-ROC, its improvements over all five baselines—including the strongest convolutional models, Deep4Net and ChronoNet—were statistically significant ($p < 0.05$). On balanced accuracy and AUC-PR the margins were narrower: Deep4Net was marginally higher on both metrics, so we do not claim an advantage on these two. On TUEP, DiffSA-EEG achieved the highest values across all six metrics (94.7% accuracy, 93.2% specificity, 98.3% AUC-ROC). It significantly outperformed EEGNet, Deep4Net, ChronoNet, and TCN ($p < 0.05$); it also surpassed the strong Diff-E backbone on every metric, although these differences did not reach statistical significance. Notably, EEGNet exhibited substantial performance degradation on TUEP (58.1% accuracy), likely due to the dataset's class imbalance and complex epileptiform patterns, whereas DiffSA-EEG maintained robust performance.

3.2. Ablation Study

We systematically evaluated all $2^4 = 16$ combinations of the four modules (SF, SDA, CBAM, self-attention) added to the Diff-E backbone (Table 2). A key observation is a *non-monotonic relationship* between architectural complexity and performance: on the two datasets studied, task-specific two-component configurations tended to outperform both simpler and fully integrated models.

On TUAB, the **SF + SDA** combination achieved the best results (84.6% accuracy, 90.8% AUC-ROC), outperforming even the full four-module configuration (83.9% accuracy). This indicates that spatial dimensionality reduction and noise robustness are most critical for heterogeneous clinical abnormalities, while the addition of attention mechanisms introduces unnecessary redundancy. On TUEP, the **SDA + CBAM** configuration yielded superior performance (96.0% accuracy, 98.4% AUC-ROC), surpassing the full model by 1.3% in accuracy. This suggests that attention-guided refinement is more effective for capturing the transient, spatially structured epileptiform discharges characteristic of TUEP.

We interpret this dataset dependence in terms of where the discriminative signal resides. TUAB abnormalities (e.g., diffuse slowing in stroke, dementia, or encephalopathy) are spatially broad and heavily artifact-contaminated, so channel-level dimensionality reduction (SF) and noise-invariant encoding (SDA) remove nuisance variation that would otherwise dominate; once this is achieved, additional attention adds parameters without a clear target and can dilute the signal, consistent with the full model under-performing SF + SDA. TUEP epileptiform activity, by contrast, is comparatively localized and transient, so the channel- and time-attention in CBAM can emphasize the relevant montage regions and time windows, while aggressive spatial compression risks discarding the focal cues. This view is consistent with the Grad-CAM analysis below, and it implies that component selection should be guided by the spatial and temporal structure of the target pathology rather than by model size alone.

3.3. Interpretability via Grad-CAM

To assess clinical plausibility, we conducted a Grad-CAM-based relevance analysis (Selvaraju et al., 2017) by projecting component-level importance back to the original scalp montage via the SVD spatial bases used during the forward spatial filtering process (Figure 2). For TUAB, the model attributed higher importance to centro-posterior and temporo-occipital regions for abnormal EEG detection, consistent with posterior slowing patterns characteristic of diffuse cerebral dysfunction. For TUEP, increased relevance was observed in right temporal and inferior-lateral regions, aligning with the well-documented temporal lobe involvement in epileptic activity. These cohort-specific patterns suggest that DiffSA-EEG captures spatial structure rather than relying on a single shared topographic signature; we note, however, that Grad-CAM provides only limited causal insight, so these observations should be read as qualitative support rather than definitive evidence.

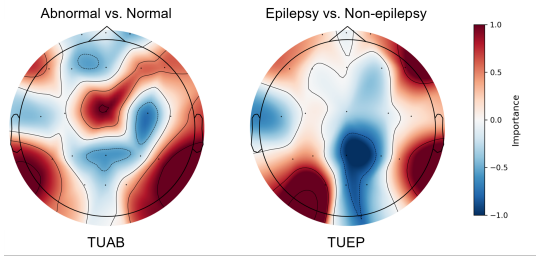

Figure 2. Average difference in Grad-CAM scalp distributions. Left: abnormal vs. normal (TUAB). Right: epilepsy vs. non-epilepsy (TUEP). Positive values indicate higher model-attributed importance for the target class.

4. Discussion and Conclusion

We presented DiffSA-EEG, a diffusion-based framework for clinical EEG classification that integrates spatial filtering, denoising, and attention mechanisms within an end-to-end pipeline. While the individual building blocks—denoising diffusion models, denoising autoencoders, attention, and SVD-initialized spatial filtering—are themselves established techniques, our contribution lies in integrating them around diffusion-based auxiliary denoising and in the empirical observation that, on the datasets studied, their most effective combination is dataset-dependent rather than monolithic. Our work offers three key insights for the structured health data community.

First, *diffusion as auxiliary denoising*: rather than using DDPMs for data generation, DiffSA-EEG trains the diffusion backbone as an auxiliary denoising task whose representations support the discriminative branch. This may be helpful for clinical EEG, which is often contaminated with heterogeneous artifacts and non-stationary noise. The framework also injects stochasticity at two different levels—a diffusion-scheduled perturbation ϵ at the raw-signal level and a fixed Gaussian perturbation ($\sigma = 0.05$) at the latent-feature level—which are intended to play complementary rather than redundant roles. We do not isolate the contribution of the diffusion component here; quantifying it against simpler denoising or noise-injection baselines is left to future work.

Second, *task-specific modularity*: the most effective module subset varied with the target pathology rather than growing with model size, suggesting that selective, dataset-aware integration may be preferable to monolithic design. This has practical implications for deploying EEG models in resource-limited clinical settings, although confirming the trend would require more than the two datasets examined here.

Third, *clinical interpretability*: the Grad-CAM topographies are broadly consistent with established neurophysiological biomarkers, offering qualitative support for the model's potential as an explainable diagnostic aid; a quantitative comparison against canonical clinical distributions, along with complementary attribution methods, remains future work.

Limitations include the need for multi-site validation to fully assess clinical readiness, and the computational overhead of diffusion-based processing. In addition, because k-means SMOTE was applied to the full TUEP set before the train/test partition, the TUEP evaluation set contains synthetic samples; the TUEP results should therefore be interpreted with this caveat, and re-evaluation on a natural-distribution, SMOTE-free test set is left for future verification. Future work will explore lightweight diffusion variants for real-time deployment and extend the framework to additional structured clinical time-series modalities. We further plan to broaden evaluation to more diverse corpora and to recent transformer- and state-space baselines, to report a complete computational-cost analysis (parameter count, training time, and inference latency relative to the Diff-E backbone), and to consolidate the statistical comparisons using per-run resampling across all evaluation metrics.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under Grant Nos. RS-2025-24533826 and RS-2025-25432692. This work was also supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) under the Information Technology Research Center (ITRC) program (Grant No. IITP-2026-RS-2020-II201808), and by an IITP grant funded by the Korean government (MSIT) (Grant No. RS-2025-02283048). This work was also supported by the Korea Ministry of Science and ICT's Special Account for Regional Balanced Development for Commercialization, supervised by the National IT Industry Promotion Agency (NIPA), to support digital medical devices for AI-based neurodevelopmental disorders (H0501-25-1001, H.S.H.).

References

Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K.-H., and Suri, J. S. Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control*, 7(4):401–408, 2012.

Gemein, L. A., Schirrmeister, R. T., Chrabąszcz, P., Wilson, D., Boedecker, J., Schulze-Bonhage, A., Hutter, F., and Ball, T. Machine-learning-based diagnostics of EEG pathology. *NeuroImage*, 220:117021, 2020.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion prob-

abilistic models. In *Advances in Neural Information Processing Systems*, pp. 6840–6851, 2020.

Jiang, Z., Dai, W., Wei, Q., Qin, Z., Li, K., and Zhang, L. EEG-DIF: Early warning of epileptic seizures through generative diffusion model-based multi-channel EEG signals forecasting. *arXiv preprint arXiv:2410.17343*, 2024.

Kim, S., Lee, Y.-E., Lee, S.-H., and Lee, S.-W. Diff-E: Diffusion-based learning for decoding imagined speech EEG. *arXiv preprint arXiv:2307.14389*, 2023.

Last, F., Douzas, G., and Bacao, F. Oversampling for imbalanced learning based on K-means and SMOTE. *Information Sciences*, 465:1–20, 2018.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15 (5):056013, 2018.

Obeid, I. and Picone, J. The Temple University Hospital EEG data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

Rommel, C., Moreau, T., Paillard, J., and Gramfort, A. CADDA: Class-wise automatic differentiable data augmentation for EEG signals. *arXiv preprint arXiv:2106.13695*, 2022.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Roy, S., Kiral-Kornek, I., and Harrer, S. ChronoNet: A deep recurrent neural network for abnormal EEG identification. In *Artificial Intelligence in Medicine*, pp. 47–56. Springer, 2019a.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. Deep learning-based electroencephalography analysis: a systematic review. volume 16, pp. 051001. IOP Publishing, 2019b.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., and Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

Shu, K., Wu, L., Zhao, Y., Liu, A., Qian, R., and Chen, X. Data augmentation for seizure prediction with generative diffusion model. *IEEE Transactions on Cognitive and Developmental Systems*, 17(3):577–591, 2025.

Tegon, A., Ingolfsson, T. M., Wang, X., Benini, L., and Li, Y. FEMBA: Efficient and scalable EEG analysis with a bidirectional Mamba foundation model. *arXiv preprint arXiv:2502.06438*, 2025.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11:3371–3408, 2010.

Wang, Y. et al. DiffMDD: A diffusion-based deep learning framework for MDD diagnosis using EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:728–738, 2024.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. CBAM: Convolutional block attention module. In *European Conference on Computer Vision*, pp. 3–19. Springer, 2018.