MultiDiffNet: A Multi-Objective Diffusion Framework for Generalizable Brain Decoding

Mengchun Zhang[*]
University of Pittsburgh
Pittsburgh, PA, USA
mengchuz@andrew.cmu.edu

Kateryna Shapovalenko[*]
Carnegie Mellon University
Pittsburgh, PA, USA
kshapova@alumni.cmu.edu

Yucheng Shao
Carnegie Mellon University
Pittsburgh, PA, USA
yshao3@andrew.cmu.edu

Eddie Guo
Carnegie Mellon University
Pittsburgh, PA, USA
yuzhiguo@andrew.cmu.edu

Parusha Pradhan
University of Pittsburgh
Pittsburgh, PA, USA
pap203@pitt.edu

Abstract

Neural decoding from electroencephalography (EEG) remains fundamentally limited by poor generalization to unseen subjects, driven by high inter-subject variability and the lack of large-scale datasets to model it effectively. Existing methods often rely on synthetic subject generation or simplistic data augmentation, but these strategies fail to scale or generalize reliably. We introduce *MultiDiffNet*, a diffusion-based framework that bypasses generative augmentation entirely by learning a compact latent space optimized for multiple objectives. We decode directly from this space and achieve state-of-the-art generalization across various neural decoding tasks using subject and session disjoint evaluation. We also curate and release a unified benchmark suite spanning four EEG decoding tasks of increasing complexity (SSVEP, Motor Imagery, P300, and Imagined Speech) and an evaluation protocol that addresses inconsistent split practices in prior EEG research. Finally, we develop a statistical reporting framework tailored for low-trial EEG settings. Our work provides a reproducible and open-source foundation for subject-agnostic EEG decoding in real-world BCI systems.

1 Introduction

Electroencephalography (EEG) is a widely used modality in brain–computer interfaces (BCIs), supporting applications from assistive communication to cognitive monitoring. Deep learning has improved decoding across motor imagery, SSVEP, and speech tasks [8, 1, 18], yet generalizing to unseen subjects remains challenging due to high inter-subject variability and limited data [12, 3].

Subject-specific models require extensive per-user calibration [9, 22], while multi-subject models struggle to generalize [25, 20, 32]. The alternative is to use two-stage pipelines that generate EEG via GANs or diffusion and then train decoders [9, 29], but they suffer from low realism, artifact transfer, and inefficiencies.

We propose *MultiDiffNet*, a unified multi-objective diffusion framework that learns a shared latent space, eliminating the need for synthetic augmentation and enhancing generalization. To benchmark progress, we release a curated suite spanning SSVEP, Motor Imagery, P300, and Imagined Speech

[*] Equal contribution.

[*] Project code: https://github.com/eddieguo-1128/DualDiff

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

tasks, with standardized subject- and session-disjoint evaluation. We also develop a statistical reporting protocol tailored for low-trial EEG research, addressing a persistent gap in reproducibility.

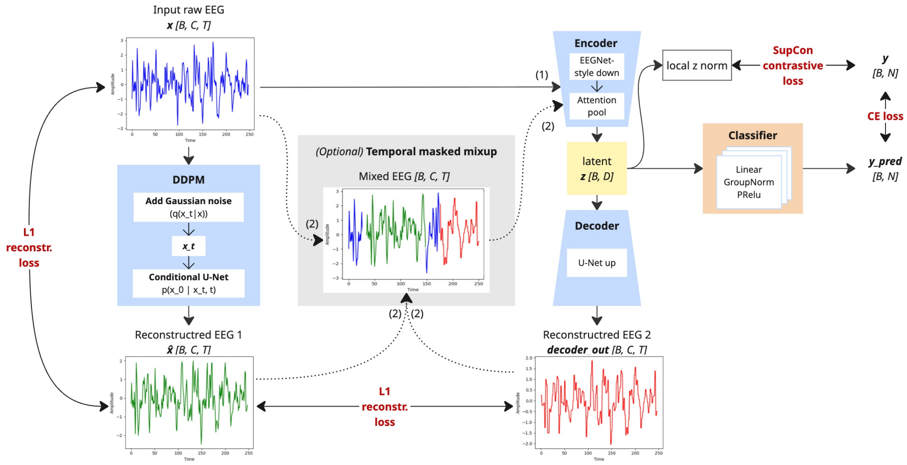


Figure 1: Overview of the *MultiDiffNet* that jointly optimizes a conditional DDPM, a contrastive encoder, and a generative decoder through a shared latent space z. The encoder produces discriminative features used for both classification and contrastive learning, while the decoder and DDPM reconstruct the input signal. An optional *temporal masked mixup* module stochastically blends the original, DDPM-denoised, and decoder-reconstructed EEG to improve representation quality.

2 Related work

We cite foundational and recent works across EEG decoding [17, 19, 26, 27, 11, 12, 3, 25, 20, 32], diffusion modeling [30, 10, 2, 29, 14, 5, 7], augmentation strategies [23, 21, 13, 24, 33], and evaluation protocols [6, 4, 36, 16, 34]. See Appendix for detailed discussion.

3 Methodology

3.1 MultiDiffNet architecture

MultiDiffNet is a modular architecture designed to jointly optimize classification, reconstruction, and contrastive structure learning from EEG signals. It consists of a Denoising Diffusion Probabilistic Model (DDPM), a discriminative encoder, a generative decoder, and a classifier (Figure 1).

Given a raw EEG signal $x \in \mathbb{R}^{C \times T}$, where C is the number of EEG channels and T is the number of timepoints, the model processes the input in two parallel paths. First, the DDPM denoises the signal via a learned reverse diffusion process, producing a refined version $\hat{x} \in \mathbb{R}^{C \times T}$. Simultaneously, the same input x is passed through an EEGNet-based encoder (See Section 3.2) to extract a latent representation $z \in \mathbb{R}^{D}$, where D is the embedding dimension. The latent vector z is then used for two purposes: (1) it is passed to a lightweight decoder to reconstruct the denoised signal $\hat{x}$, resulting in a reconstruction $x_{\text{dec}} \in \mathbb{R}^{C \times T}$; and (2) it is passed to a fully connected classification head to predict class logits $\hat{y} \in \mathbb{R}^{K}$, where K is the number of classes.

To further structure the latent space, z is locally normalized (Section 3.3) and then projected to $z_{\text{proj}} \in \mathbb{R}^{D'}$, which is optimized with a supervised contrastive loss. All classification and reconstruction are performed directly from z, without relying on generated augmentations.

We performed an extensive ablation study across architectural variants, modifying the presence of DDPMs, encoder inputs, decoder pathways, classifier heads, and loss terms. The configuration described here reflects the best-performing combination.

3.2 EEGNet-style encoder with attention pool

Given EEGNet's demonstrated effectiveness across multiple EEG decoding tasks, we adapt its architecture as our discriminative encoder, hypothesizing that its proven feature extraction capabilities can produce powerful latent representations z for our multi-objective framework. Our encoder extracts multi-scale features (dn_1, dn_2, dn_3) from different layers and applies attention pooling:

$$z = \text{AttentionPool}(dn_3) \in \mathbb{R}^D,$$

3.3 Subject-wise latent normalization

To mitigate inter-subject variability, we apply subject-wise normalization on the encoder output z:

$$z_{\text{norm}} = \frac{z - \mu_s}{\sigma_s},$$

where μ_s and σ_s denote the mean and standard deviation computed per subject s using a subset of training trials. During evaluation, we adopt a two-mode strategy: for seen subjects, normalization uses pre-computed statistics from their training data; for unseen subjects, statistics are estimated on-the-fly using their own calibration trials, simulating realistic deployment scenarios.

3.4 Mixup strategies

Mixup strategies can improve robustness in low-trial EEG decoding. However, standard mixup techniques may not fully exploit the structure of neural time series. We therefore explore two complementary strategies: *Weighted Average Mixup* and a novel *Temporal Masked Mixup*. *Weighted Average Mixup* performs linear interpolation between the original EEG input x, the DDPM-denoised output $\hat{x}$, and the decoder reconstruction x_{dec}. We investigate multiple integration points in the model: **(0)** Input-level mixup, **(1-3)** Mixup after encoder layers 1, 2, or 3, respectively, **(4)** Mixup after the final attention pooling layer. To address the limitations of global interpolation, we propose *Temporal Masked Mixup*, which perturbs only localized segments of the input time series while preserving surrounding structure. See Appendix for pseudocode and Figure 1 for an illustration.

3.5 Loss functions

MultiDiffNet is trained using a weighted sum of three objectives:

$$\mathcal{L}_{\text{total}} = \underbrace{\alpha\, \mathcal{L}_{\text{CE/MSE}}(\hat{y}, y)}_{\text{classification}} + \underbrace{\beta\, \mathcal{L}_{\text{L1}}(x_{\text{dec}}, \hat{x})}_{\text{reconstruction}} + \underbrace{\gamma\, \mathcal{L}_{\text{SupCon}}(z_{\text{proj}}, y)}_{\text{contrastive}}$$

We fix $\alpha = 1.0$ and progressively scale β and γ to stabilize training:

$$\beta = \min\left(1.0, \frac{\text{epoch}}{100}\right) \cdot 0.05, \quad \gamma = \min\left(1.0, \frac{\text{epoch}}{50}\right) \cdot 0.2$$

Details on loss formulation and weighting strategies are provided in the Appendix.

3.6 Evaluation metrics

We evaluate model performance primarily using downstream classification accuracy, which quantifies the proportion of correctly classified EEG samples. Accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

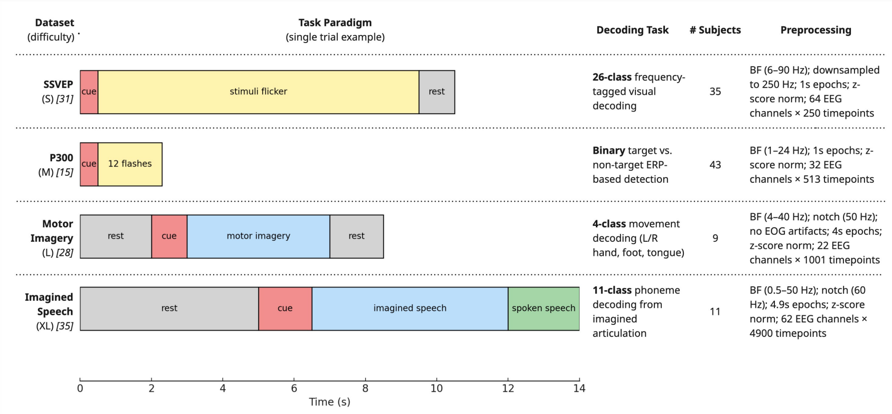
where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. In addition, we report F1 score, precision, recall, and AUC for a more comprehensive evaluation; detailed formulas and results are provided in the Appendix.

3.7 Trend-level statistical reporting framework

Conventional p-values often fail under the high-variance, low-trial, subject-disjoint conditions of EEG decoding. To address this, we introduce a robust trend-level statistical framework (detailed in the Appendix) that synthesizes effect sizes, cross-seed consistency, and Bayesian posterior probabilities. This allows us to detect systematic, reproducible gains even when classical significance tests return null results. Our approach represents a principled shift toward reproducible, evidence-based model evaluation in brain decoding.

4 Experiments and results

4.1 Benchmark dataset suite



Dataset (difficulty)	Decoding Task	# Subjects	Preprocessing
SSVEP (S) *[31]*	**26-class** frequency-tagged visual decoding	35	BF (6–90 Hz); downsampled to 250 Hz; 1s epochs; z-score norm; 64 EEG channels × 250 timepoints
P300 (M) *[15]*	**Binary** target vs. non-target ERP-based detection	43	BF (1–24 Hz); 1s epochs; z-score norm; 32 EEG channels × 513 timepoints
Motor Imagery (L) *[28]*	**4-class** movement decoding (L/R hand, foot, tongue)	9	BF (4–40 Hz); notch (50 Hz); no EOG artifacts; 4s epochs; z-score norm; 22 EEG channels × 1001 timepoints
Imagined Speech (XL) *[35]*	**11-class** phoneme decoding from imagined articulation	11	BF (0.5–50 Hz); notch (60 Hz); 4.9s epochs; z-score norm; 62 EEG channels × 4900 timepoints

Figure 2: Overview of four EEG datasets ranked by task difficulty from easiest (top) to hardest (bottom). Task paradigms and preprocessing details are adapted from the original publications: SSVEP [31], P300 [15], Motor Imagery [28], and Imagined Speech [35].

We curated four diverse EEG benchmarks (SSVEP, P300, Motor Imagery, and Imagined Speech), spanning increasing decoding difficulty. Each dataset is split into train, val, and two test sets: a seen-subject (intra-subject) split and an unseen-subject (cross-subject) split. This standardized protocol enables rigorous evaluation of both personalization and generalization, addressing the inconsistent and often unrealistic split practices prevalent in prior EEG research, where models are evaluated on mixed subject data or using computationally expensive LOSO.

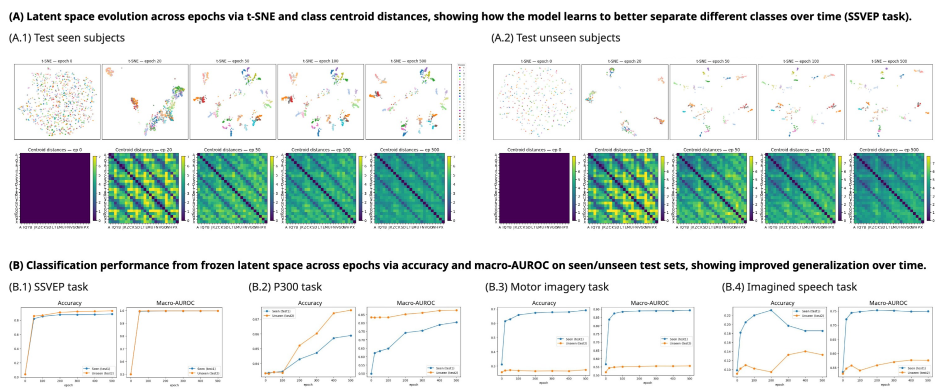
4.2 Generalization performance

MultiDiffNet delivers a decisive leap in generalization. Unlike raw EEG representations, where class boundaries blur due to subject-specific noise, our learned latent space forms clearly separable, label-aligned clusters (Figure 3). This structured representation enables robust decoding across subjects. As shown in Table 1, *MultiDiffNet* consistently reduces the seen–unseen accuracy gap across all tasks. In SSVEP, it lifts cross-subject accuracy from 81.08% (EEGNet) to 84.72%, further boosted to 85.25% with Temporal Masked Mixup. Even in the low-SNR regime of Imagined Speech, it delivers a +6.3% gain over baseline. While effects on P300 and Motor Imagery are smaller—likely due to ceiling effects or dataset limitations, our results decisively demonstrate that *MultiDiffNet* learns invariant, generalizable representations across a wide spectrum of EEG decoding challenges.

4.3 Ablation studies

To understand what drives generalization in *MultiDiffNet*, we ran extensive ablation experiments, over 100 controlled configs. All results are reported for both seen- and unseen-subject accuracy, with statistical evidence matrices and trend-level effect sizes in the Appendix.

Decoder input. Feeding only z to the decoder often matches or exceeds more complex fusion variants. For example, SSVEP unseen accuracy reaches 84.72% with z alone, further boosted to 85.25%


(A) Latent space evolution across epochs via t-SNE and class centroid distances, showing how the model learns to better separate different classes over time (SSVEP task).

(B) Classification performance from frozen latent space across epochs via accuracy and macro-AUROC on seen/unseen test sets, showing improved generalization over time.

Figure 3: (A) Visualization of latent space across training epochs. (B) Downstream classification performance from frozen latent representations.

Table 1: Final results across tasks and models. Accuracy is reported for both seen-subject (intra-subject) and unseen-subject (cross-subject) test splits. Tasks are ranked by task difficulty. *Stars denote win percentage:* *** $\geq$ 80%, ** $\geq$ 60%, * $\geq$ 40%. Detailed results are in the Appendix.

Task	**Model**	**Subj.**	**Classes**	**Seen Acc. (%)**	**Unseen Acc. (%)**
SSVEP	EEGNet	35	26	$89.16 \pm 0.57^{***}$	$81.08 \pm 9.16^{*}$
	EEGNet + Vanilla Aug.	35	26	74.01 ± 1.43	30.02 ± 3.16
	MultiDiffNet	35	26	$85.08 \pm 1.53^{*}$	$\mathbf{84.72 \pm 6.03}^{**}$
	MultiDiffNet + Temp. Mixup	35	26	$86.79 \pm 1.75^{**}$	$\mathbf{85.25 \pm 6.94}^{***}$
P300	EEGNet	43	2	$88.79 \pm 0.67^{**}$	$87.24 \pm 2.01^{**}$
	MultiDiffNet	43	2	85.35 ± 1.12	$79.47 \pm 0.54^{*}$
	MultiDiffNet + Temp. Mixup	43	2	85.61 ± 0.52	79.56 ± 4.43
MI	EEGNet	9	4	$67.01 \pm 5.38^{**}$	$46.18 \pm 7.20^{**}$
	MultiDiffNet	9	4	55.85 ± 2.80	39.24 ± 8.00
	MultiDiffNet + Temp. Mixup	9	4	$57.69 \pm 3.27^{*}$	36.78 ± 5.23
Img. Speech	EEGNet	14	11	11.26 ± 2.01	10.61 ± 0.93
	MultiDiffNet	14	11	$\mathbf{15.55 \pm 0.62}^{*}$	$\mathbf{11.62 \pm 1.29}^{*}$
	MultiDiffNet + Temp. Mixup	14	11	$\mathbf{17.57 \pm 1.16}^{**}$	$\mathbf{12.12 \pm 0.38}^{**}$

with mixup, while more elaborate fusions ($z + x$, x_{hat} + skips) show no consistent gains. These findings validate our architectural decision to decode primarily from z.

Classifier head. A lightweight FC head on z delivers state-of-the-art generalization with minimal complexity. It rivals or outperforms EEGNet classifiers trained on x, especially in low-SNR tasks. This supports our choice to use FC as the default classification head.

Encoder and decoder. Using raw x as encoder input consistently outperforms $\hat{x}$, showing that denoising is useful for regularization. Interestingly, removing the decoder entirely sometimes improves generalization, suggesting that reconstruction may introduce noise if overemphasized.

Loss combinations. Combining CE with mild MSE or contrastive losses improves stability, particularly when auxiliary weights are gently annealed. The best results use $\beta = 0.05$, $\gamma = 0.2$—balancing reconstruction as a regularizer without overpowering the classification objective.

Mixup strategies. Mixup effects are task-specific. For SSVEP, *Temporal Masked Mixup* outperforms all variants. Motor Imagery benefits from *Weighted Average Mixup*, while P300 and Imagined Speech show limited sensitivity, highlighting that mixup is most impactful in high-SNR regimes.

5 Conclusions and future work

We presented *MultiDiffNet*, a diffusion-based neural decoder that learns a compact, multi-objective latent space for EEG decoding without synthetic augmentation. Through unified benchmarks and rigorous cross-subject evaluation, we showed that *MultiDiffNet* achieves strong generalization across diverse BCI paradigms, particularly in challenging low-signal settings such as SSVEP and Imagined Speech. Our statistical analysis framework further addresses reproducibility challenges in low-trial EEG research. Future work will explore scaling *MultiDiffNet* to larger and more diverse EEG datasets and extending the architecture to other neural modalities.

Acknowledgments

The authors would like to thank Professor Bhiksha Raj of Carnegie Mellon University for his guidance and support.

References

[1] H. Ahmadi and L. Mesin. Universal semantic feature extraction from eeg signals: a task-independent framework. *Journal of Neural Engineering*, 22(3):036003, 2025.

[2] Y. An, Y. Tong, W. Wang, and S. W. Su. Enhancing eeg signal generation through a hybrid approach integrating reinforcement learning and diffusion models, 2024. URL https://arxiv.org/abs/2410.00013.

[3] K. Barmpas, Y. Panagakis, S. Bakas, D. A. Adamos, N. Laskaris, and S. Zafeiriou. Improving generalization of cnn-based motor-imagery eeg decoders via dynamic convolutions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1997–2005, 2023.

[4] Z. Chen, P. T. Wang, M. Ibrahim, S. Baveja, R. Mu, A. H. Do, and Z. Nenadic. Leveraging transfer learning and user-specific updates for rapid training of bci decoders. *arXiv preprint arXiv:2506.14120*, 2025.

[5] W. Chow, J. Li, Q. Yu, K. Pan, H. Fei, Z. Ge, S. Yang, S. Tang, H. Zhang, and Q. Sun. Unified generative and discriminative training for multi-modal large language models. *Advances in Neural Information Processing Systems*, 37:23155–23190, 2024.

[6] F. Del Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, L. Finos, and M. Atzori. The role of data partitioning on the performance of eeg-based deep learning models in supervised cross-subject analysis: a preliminary study. *Computers in Biology and Medicine*, 196:110608, 2025.

[7] W. Grathwohl, K.-C. Wang, J.-H. Jacobsen, D. Duvenaud, M. Norouzi, and K. Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

[8] H. Gu, T. Chen, X. Ma, M. Zhang, Y. Sun, and J. Zhao. Cltnet: A hybrid deep learning model for motor imagery classification. *Brain Sciences*, 15(2):124, 2025.

[9] K. G. Hartmann, R. T. Schirrmeister, and T. Ball. Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals. *arXiv preprint arXiv:1806.01875*, 2018.

[10] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[11] W. Hu, G. Jiang, J. Han, X. Li, and P. Xie. Regional-asymmetric adaptive graph convolutional neural network for diagnosis of autism in children with resting-state eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:200–211, 2023.

[12] G. Huang, Z. Zhao, S. Zhang, Z. Hu, J. Fan, M. Fu, J. Chen, Y. Xiao, J. Wang, and G. Dan. Discrepancy between inter-and intra-subject variability in eeg-based motor imagery brain-computer interface: Evidence from multiple perspectives. *Frontiers in neuroscience*, 17: 1122661, 2023.

[13] G. Kim, D. K. Han, and H. Ko. Specmix: A mixed sample data augmentation method for training withtime-frequency domain features. *arXiv preprint arXiv:2108.03020*, 2021.

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

[15] L. Korczowski, M. Cederhout, A. Andreev, G. Cattan, P. L. C. Rodrigues, V. Gautheret, and M. Congedo. *Brain Invaders calibration-less P300-based BCI with modulation of flash duration Dataset (bi2015a)*. PhD thesis, GIPSA-lab, 2019.

[16] S. Kunjan, T. S. Grummett, K. J. Pope, D. M. Powers, S. P. Fitzgibbon, T. Bastiampillai, M. Battersby, and T. W. Lewis. The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis. In *International conference on brain informatics*, pages 558–567. Springer, 2021.

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

[18] Y.-E. Lee and S.-H. Lee. Eeg-transformer: Self-attention from transformer architecture for decoding eeg of imagined speech. In *2022 10th International winter conference on brain-computer interface (BCI)*, pages 1–4. IEEE, 2022.

[19] W. Liao, H. Liu, and W. Wang. Advancing bci with a transformer-based model for motor imagery classification. *Scientific Reports*, 15(1):23380, 2025.

[20] S. Liu, J. Zhang, A. Wang, H. Wu, Q. Zhao, and J. Long. Subject adaptation convolutional neural network for eeg-based motor imagery classification. *Journal of Neural Engineering*, 19 (6):066003, 2022.

[21] X.-H. Liu, B.-L. Lu, and W.-L. Zheng. mixeeg: Enhancing eeg federated learning for cross-subject eeg classification with tailored mixup. *arXiv preprint arXiv:2504.07987*, 2025.

[22] T.-j. Luo and Z. Cai. Diffusion models-based motor imagery eeg sample augmentation via mixup strategy. *Expert Systems with Applications*, 235:125585, 2024. doi: 10.1016/j.eswa.2024.125585. URL https://www.sciencedirect.com/science/article/pii/S0957417424024527.

[23] T.-j. Luo and Z. Cai. Diffusion models-based motor imagery eeg sample augmentation via mixup strategy. *Expert Systems with Applications*, 262:125585, 2025.

[24] Y. Pei, Z. Luo, Y. Yan, H. Yan, J. Jiang, W. Li, L. Xie, and E. Yin. Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg. *Frontiers in Human Neuroscience*, 15:645952, 2021.

[25] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort. Data augmentation for learning predictive models on eeg: a systematic comparison. *Journal of Neural Engineering*, 19(6):066020, 2022.

[26] Y. Song, Q. Zheng, B. Liu, and X. Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

[27] X. Tang, J. Zhang, Y. Qi, K. Liu, R. Li, and H. Wang. A spatial filter temporal graph convolutional network for decoding motor imagery eeg signals. *Expert Systems with Applications*, 238:121915, 2024.

[28] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

[29] S. Torma and L. Szegletes. Generative modeling and augmentation of eeg signals using improved diffusion probabilistic models. *Journal of Neural Engineering*, 22(1):016001, 2025. doi: 10.1088/1741-2552/ada0e4.

[30] G. Tosato, C. M. Dalbagno, and F. Fumagalli. Eeg synthetic data generation using probabilistic diffusion models, 2023. URL https://arxiv.org/abs/2303.06068.

[31] Y. Wang, X. Chen, X. Gao, and S. Gao. A benchmark dataset for ssvep-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10): 1746–1752, 2017. doi: 10.1109/TNSRE.2016.2578980. URL https://ieeexplore.ieee.org/document/7740878.

[32] D. Wu. Online and offline domain adaptation for reducing bci calibration effort. *IEEE Transactions on human-machine Systems*, 47(4):550–563, 2016.

[33] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

[34] H. Zhang, H. Ji, J. Yu, J. Li, L. Jin, L. Liu, Z. Bai, and C. Ye. Subject-independent eeg classification based on a hybrid neural network. *Frontiers in Neuroscience*, 17:1124089, 2023.

[35] S. Zhao and F. Rudzicz. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 992–996. IEEE, 2015.

[36] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng. Ctnet: a convolutional transformer network for eeg-based motor imagery classification. *Scientific reports*, 14(1):20237, 2024.