MULTIDIFFNET: A MULTI-OBJECTIVE DIFFUSION FRAMEWORK FOR GENERALIZABLE BRAIN DECODING

Anonymous authors
Paper under double-blind review

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 Gu et al. (2025); Ahmadi & Mesin (2025); Lee & Lee (2022), yet generalizing to unseen subjects remains challenging due to high inter-subject variability and limited data Huang et al. (2023); Barmpas et al. (2023).

Subject-specific models require extensive per-user calibration Hartmann et al. (2018); Luo & Cai (2024), while multi-subject models struggle to generalize Rommel et al. (2022); Liu et al. (2022); Wu (2016). The alternative is to use two-stage pipelines that generate EEG via GANs or diffusion and then train decoders (Hartmann et al., 2018; Torma & Szegletes, 2025), 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 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.

2 RELATED WORK

EEG Decoding and Generalization EEG decoding has evolved from handcrafted features to deep architectures, with EEGNet emerging as a widely adopted baseline due to its efficient depthwise–separable convolutions and lightweight design (Lawhern et al., 2018). Recent models explore transformers (Liao et al., 2025; Song et al., 2022a) and graph neural networks (Tang et al., 2024; Hu et al., 2023), but EEGNet remains favored for its robustness and simplicity. A key limitation is poor cross-subject generalization, with 20-40% accuracy drops despite strong within-subject performance (Huang et al., 2023; Barmpas et al., 2023). Attempts to address this require expensive calibration (Rommel et al., 2022; Liu et al., 2022; Wu, 2016). Scalable BCIs require subject-agnostic models that generalize without per-user retraining.

Diffusion Models for EEG Denoising Diffusion Probabilistic Models (DDPMs) model data distributions via iterative denoising and outperform GANs in EEG synthesis by avoiding mode collapse (Tosato et al., 2023; Ho et al., 2020). Recent enhancements, such as reinforcement learning (An et al., 2024) and progressive distillation (Torma & Szegletes, 2025), have further improved realism and sampling speed. Diff-E (Kim et al., 2023) extended diffusion to imagined-speech decoding via joint reconstruction and classification, but remained task-specific and did not address cross-subject generalization. Broader research suggests that combining generative and discriminative objectives yields stronger representations (Chow et al., 2024; Grathwohl et al., 2019), yet EEG models typically optimize only one. We explore this joint learning paradigm across diverse EEG tasks, aiming to learn generalizable representations that capture both signal structure and task-relevant information.

Mixup Methods Signal-level augmentation has evolved from basic jittering and filtering to temporal, spectral, and channel-wise mixup (Luo & Cai, 2025; Liu et al., 2025; Kim et al., 2021; Pei et al., 2021; Zhang et al., 2017), but many variants introduce unrealistic artifacts that hinder generalization. This motivates our systematic evaluation of weighted and temporal input mixup across encoder layers, along with latent-space mixing

Evaluation Strategies Effective cross-subject EEG decoding requires both rigorous training strategies and standardized evaluation. Leave-one-subject-out (LOSO) validation remains common but is computationally intensive and impractical for real-time deployment (Del Pup et al., 2025; Chen et al., 2025; Zhao et al., 2024; Barmpas et al., 2023; Kunjan et al., 2021), while simpler subject splits often neglect session independence and true seen/unseen separation (Zhang et al., 2023). We address it in our work by introducing a standardized subject- and session-disjoint evaluation.

3 METHODOLOGY

3.1 MULTIDIFFNET ARCHITECTURE

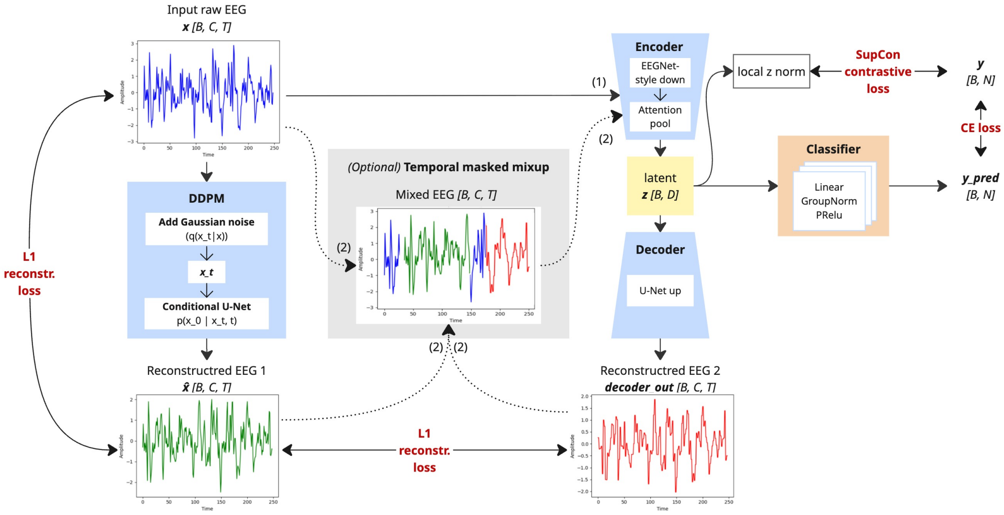


Figure 1: Overview of the *MultiDiffNet* that jointly optimizes a conditional DDPM, a contrastive encoder, and a generative decoder through a shared latent space z. The encoder produces discriminative features used for both classification and contrastive learning, while the decoder and DDPM reconstruct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

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 Algorithm 1 for pseudocode.

3.5 Loss functions

MultiDiffNet is trained using a weighted sum of three objectives:

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

Algorithm 1 Temporal Masked Mixup

```
Initialize a binary mask $M \in \{0,1\}^{C \times T}$ with all zeros.
Flip each 0 in $M$ to 1 with probability $p = 0.01$.
for each position in $M$ with value 1 do
    Expand to a temporal window of random length (uniform between min and max size).
end for
Flip each 1 in $M$ to $-1$ with:
    • Fixed probability 0.5 (fixed ratio), or
    • Probability drawn from Beta$(0.2, 0.2)$ each epoch (random ratio).
Apply the final mask:
    • $0 \to x$ (original input)
    • $1 \to \hat{x}$ (DDPM output)
    • $-1 \to x_{\text{dec}}$ (decoder output)
```

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

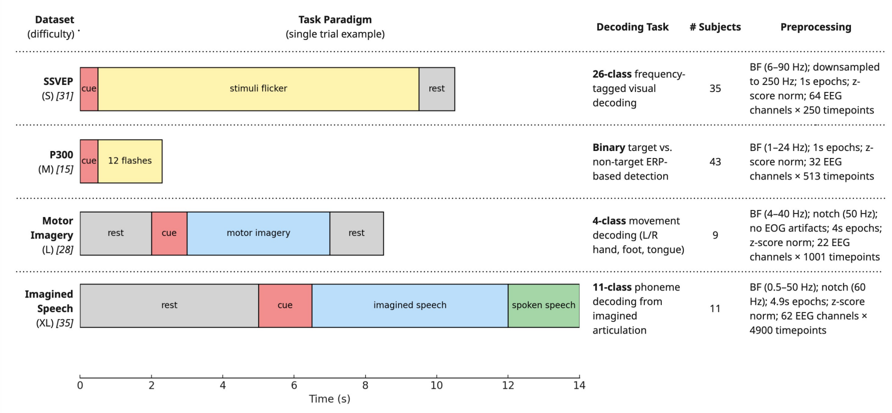
Conventional p-values often fail under the high-variance, low-trial, subject-disjoint conditions of EEG decoding. To address this, we introduce a robust trend-level statistical framework (detailed in the Appendix) that synthesizes effect sizes, cross-seed consistency, and Bayesian posterior probabilities. 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.

While this framework enhances reproducibility, it is not meant to substitute conventional p-value testing. Instead, it addresses a well-documented limitation: in low-trial, high-variance EEG decoding, even systematic improvements may fail to reach arbitrary significance thresholds. Such small yet consistent gains—for instance, 2–3% accuracy in imagined speech or SSVEP—can substantially affect usability in BCI systems. By combining effect sizes, cross-seed consistency, and Bayesian evidence, the framework provides a principled way to surface these domain-relevant improvements, while remaining fully compatible with classical and non-parametric statistical tests.

4 Experiments and results

4.1 Benchmark dataset suite

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



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 Wang et al. (2017), P300 Korczowski et al. (2019), Motor Imagery Tangermann et al. (2012), and Imagined Speech Zhao & Rudzicz (2015).

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 Baselines

We benchmarked our model against a diverse set of carefully selected baselines to ensure robust and fair comparisons. Our selection criteria were twofold: (i) prioritize architectures that are widely used for generalization to unseen subjects or sessions, and (ii) cover the main inductive biases found in EEG decoding, such as spatial filtering, temporal modeling, and attention mechanisms.

Specifically, we include: (1) **EEGNet** (Lawhern et al., 2018), a compact depthwise-separable CNN that is widely adopted for cross-subject generalization due to its strong accuracy–efficiency trade-off; (2) **ShallowFBCSPNet** (Schirrmeister et al., 2017), which implements learnable filter-bank Common Spatial Patterns (CSP) to extract frequency–spatial features; (3) **TIDNet** (Kostas & Rudzicz, 2020), which introduces dilated convolutions and residual connections to improve robustness under subject shift; (4) **EEGConformer** (Song et al., 2022b), which combines a convolutional front-end with self-attention to model both local spatial structure and global temporal context; and (5) **EEGTCNet** (Ingolfsson et al., 2020), a temporal convolutional network tailored for EEG that emphasizes causal and dilated temporal modeling, offering complementary inductive bias to purely spatial–spectral models.

All models are evaluated using identical input windows of shape (C, T), and trained with a unified global training schedule to ensure comparability. Public implementations and recommended hyperparameters are used where available, with no method-specific tuning.

4.3 Generalization performance

MultiDiffNet helps with 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 baseline) to 84.72%, further boosted to 85.25% with Temporal Masked Mixup. For comparison, other representative architectures such as ShallowFBCSPNet (58.87%), EEGConformer (51.92%), TIDNet (25.96%), and EEGTCNet (49.57%) fall well behind, highlighting the robustness of our latent-space design.

Even in the low-SNR regime of Imagined Speech, *MultiDiffNet* improves cross-subject accuracy from 10.61% (EEGNet) to 12.12%, while simultaneously achieving a much larger gain on seen-

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	ShallowFBCSPNet	35	26	69.58 ± 1.30*	58.87 ± 9.37*
	EEGConformer	35	26	66.98 ± 2.83	51.92 ± 9.06
	TIDNet	35	26	28.01 ± 4.12	25.96 ± 5.29
	EEGTCNet	35	26	58.31 ± 4.02	49.57 ± 9.14
	EEGNet	35	26	89.16 ± 0.57***	81.08 ± 9.16**
	MultiDiffNet	35	26	85.08 ± 1.53**	**84.72 ± 6.03*****
	MultiDiffNet + Mixup	35	26	86.79 ± 1.75***	**85.25 ± 6.94*****
P300	ShallowFBCSPNet	43	2	87.72 ± 0.33	86.20 ± 1.45
	EEGConformer	43	2	88.54 ± 0.54**	86.30 ± 1.73
	TIDNet	43	2	88.24 ± 0.31*	85.63 ± 0.58**
	EEGTCNet	43	2	88.69 ± 0.59***	87.02 ± 1.62***
	EEGNet	43	2	88.79 ± 0.67***	87.24 ± 2.01***
	MultiDiffNet	43	2	85.35 ± 1.12	79.47 ± 0.54*
	MultiDiffNet + Mixup	43	2	85.61 ± 0.52	79.56 ± 4.43
MI	ShallowFBCSPNet	9	4	64.34 ± 3.61***	36.46 ± 6.60
	EEGConformer	9	4	59.57 ± 5.60**	36.49 ± 7.72
	TIDNet	9	4	44.27 ± 2.60	34.42 ± 3.60
	EEGTCNet	9	4	58.85 ± 4.54	32.99 ± 6.94
	EEGNet	9	4	67.01 ± 5.38***	46.18 ± 7.20***
	MultiDiffNet	9	4	55.85 ± 2.80	39.24 ± 8.00***
	MultiDiffNet + Mixup	9	4	57.69 ± 3.27*	36.78 ± 5.23
Img. Speech	ShallowFBCSPNet	11	11	13.78 ± 1.55**	10.48 ± 0.64
	EEGConformer	11	11	10.62 ± 0.82	9.21 ± 3.00
	TIDNet	11	11	9.10 ± 0.54	10.35 ± 0.18
	EEGTCNet	11	11	12.64 ± 1.58	10.10 ± 0.64
	EEGNet	11	11	11.26 ± 2.01*	10.61 ± 0.93*
	MultiDiffNet	11	11	**15.55 ± 0.62*****	**11.62 ± 1.29*****
	MultiDiffNet + Mixup	11	11	**17.57 ± 1.16*****	**12.12 ± 0.38*****

subject accuracy (11.26% → 17.57%). Other baselines such as ShallowFBCSPNet (10.48/13.78%), EEGConformer (9.21/10.62%), TIDNet (10.35/9.10%), and EEGTCNet (10.10/12.64%) hover close to chance level on both splits, further highlighting the robustness of our approach. For such a challenging task, even modest absolute gains are meaningful, as they can indicate more reliable signal extraction under extreme noise conditions. On Motor Imagery, *MultiDiffNet* also surpasses most baselines on unseen accuracy, e.g., outperforming TIDNet (34.42%) and EEGTCNet (32.99%), while maintaining competitive seen accuracy (57.69% vs. 44.27% for TIDNet and 58.85% for EEGTCNet). Although it remains slightly below EEGNet (46.18/67.01%), this is likely due to ceiling effects and dataset scale.

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

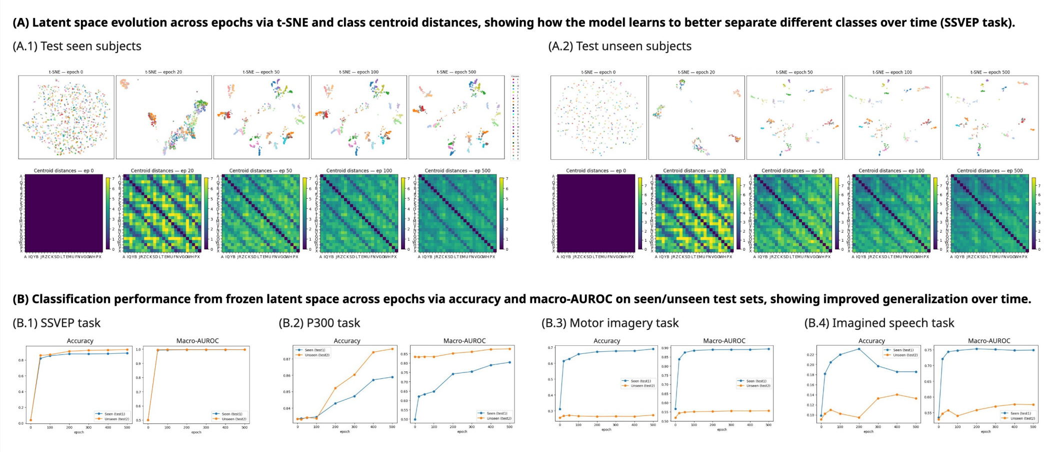


Figure 3: (A) Visualization of latent space across training epochs. (B) Downstream classification performance from frozen latent representations.

accuracies achieved in this ablation are 85.86/84.72 on SSVEP, 85.88/81.41 on P300, 56.89/40.36 on Motor Imagery, and 18.58/12.88 on Imagined Speech (seen/unseen).

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. For completeness, the best accuracies achieved in this ablation are 85.08/84.72 on SSVEP, 85.35/84.12 on P300, 55.85/39.24 on Motor Imagery, and 17.95/11.61 on Imagined Speech (seen/unseen).

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. For completeness, the best accuracies in this ablation are 90.95/85.58 on SSVEP, 85.71/80.93 on P300, 55.85/40.16 on Motor Imagery, and 19.22/13.76 on Imagined Speech (seen/unseen).

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. For completeness, the best accuracies in this ablation are 86.40/85.58 on SSVEP, 85.69/80.18 on P300, 59.67/41.44 on Motor Imagery, and 19.60/13.51 on Imagined Speech (seen/unseen).

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. For completeness, the best accuracies in this ablation are 87.84/85.26 on SSVEP, 85.78/79.56 on P300, 63.44/38.83 on Motor Imagery, and 19.47/12.12 on Imagined Speech (seen/unseen).

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.

For completeness, we note that our trend-level statistical framework is intended only as a complementary tool for low-trial EEG research; detailed rationale is provided in Section 3.7, with Bayesian and non-parametric validations reported in the Appendix.

References

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

Yang An, Yuhao Tong, Weikai Wang, and Steven W. Su. Enhancing eeg signal generation through a hybrid approach integrating reinforcement learning and diffusion models, 2024. URL https://arxiv.org/abs/2410.00013.

Konstantinos Barmpas, Yannis Panagakis, Stylianos Bakas, Dimitrios A Adamos, Nikolaos Laskaris, and Stefanos 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.

Ziheng Chen, Po T Wang, Mina Ibrahim, Shivali Baveja, Rong Mu, An H Do, and Zoran Nenadic. Leveraging transfer learning and user-specific updates for rapid training of bci decoders. *arXiv preprint arXiv:2506.14120*, 2025.

Wei Chow, Juncheng Li, Qifan Yu, Kaihang Pan, Hao Fei, Zhiqi Ge, Shuai Yang, Siliang Tang, Hanwang Zhang, and Qianru Sun. Unified generative and discriminative training for multi-modal large language models. *Advances in Neural Information Processing Systems*, 37:23155–23190, 2024.

Federico Del Pup, Andrea Zanola, Louis Fabrice Tshimanga, Alessandra Bertoldo, Livio Finos, and Manfredo 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.

Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

He Gu, Tingwei Chen, Xiao Ma, Mengyuan Zhang, Yan Sun, and Jian Zhao. Cltnet: A hybrid deep learning model for motor imagery classification. *Brain Sciences*, 15(2):124, 2025.

Kay Gregor Hartmann, Robin Tibor Schirrmeister, and Tonio Ball. Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals. *arXiv preprint arXiv:1806.01875*, 2018.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Wanyu Hu, Guoqian Jiang, Junxia Han, Xiaoli Li, and Ping 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.

Gan Huang, Zhiheng Zhao, Shaorong Zhang, Zhenxing Hu, Jiaming Fan, Meisong Fu, Jiale Chen, Yaqiong Xiao, Jun Wang, and Guo 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.

T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. arXiv preprint arXiv:2006.00622, 2020. URL https://doi.org/10.48550/arXiv.2006.00622. EEGTCNet.

Gwantae Kim, David K Han, and Hanseok Ko. Specmix: A mixed sample data augmentation method for training withtime-frequency domain features. *arXiv preprint arXiv:2108.03020*, 2021.

Soowon Kim, Young-Eun Lee, Seo-Hyun Lee, and Seong-Whan Lee. Diff-e: Diffusion-based learning for decoding imagined speech eeg. *arXiv preprint arXiv:2307.14389*, 2023.

Louis Korczowski, Martine Cederhout, Anton Andreev, Grégoire Cattan, Pedro Luiz Coelho Rodrigues, Violette Gautheret, and Marco Congedo. *Brain Invaders calibration-less P300-based BCI with modulation of flash duration Dataset (bi2015a)*. PhD thesis, GIPSA-lab, 2019.

Dimitrios Kostas and Frank Rudzicz. Thinker invariance: Enabling deep neural networks for bci across more people. *Journal of Neural Engineering*, 17(5):056008, 2020. doi: 10.1088/1741-2552/abb7a7. TIDNet.

Sajeev Kunjan, Tyler S Grummett, Kenneth J Pope, David MW Powers, Sean P Fitzgibbon, T Bastiampillai, M Battersby, and Trent W Lewis. The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis. In *International conference on brain informatics*, pp. 558–567. Springer, 2021.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Young-Eun Lee and Seo-Hyun 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)*, pp. 1–4. IEEE, 2022.

Wangdan Liao, Hongyun Liu, and Weidong Wang. Advancing bci with a transformer-based model for motor imagery classification. *Scientific Reports*, 15(1):23380, 2025.

Siwei Liu, Jia Zhang, Andong Wang, Hanrui Wu, Qibin Zhao, and Jinyi Long. Subject adaptation convolutional neural network for eeg-based motor imagery classification. *Journal of Neural Engineering*, 19(6):066003, 2022.

Xuan-Hao Liu, Bao-Liang Lu, and Wei-Long Zheng. mixeeg: Enhancing eeg federated learning for cross-subject eeg classification with tailored mixup. *arXiv preprint arXiv:2504.07987*, 2025.

Tian-jian Luo and Zikun 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.

Tian-jian Luo and Zikun Cai. Diffusion models-based motor imagery eeg sample augmentation via mixup strategy. *Expert Systems with Applications*, 262:125585, 2025.

Yu Pei, Zhiguo Luo, Ye Yan, Huijiong Yan, Jing Jiang, Weiguo Li, Liang Xie, and Erwei Yin. Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg. *Frontiers in Human Neuroscience*, 15:645952, 2021.

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.

Robin T Schirrmeister, Jost T Springenberg, Lorenz DJ Fiederer, Markus Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 2017. doi: 10.1002/hbm.23730. URL http://dx.doi.org/10.1002/hbm.23730. ShallowFBCSPNet.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022a.

Yujie Song, Qinglin Zheng, Bo Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022b. doi: 10.1109/TNSRE.2022.3223706. URL https://ieeexplore.ieee.org/document/9991178. EEGConformer.

Xianlun Tang, Jing Zhang, Yidan Qi, Ke Liu, Rui Li, and Huiming Wang. A spatial filter temporal graph convolutional network for decoding motor imagery eeg signals. *Expert Systems with Applications*, 238:121915, 2024.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot R Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

Szabolcs Torma and Luca 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.

Giulio Tosato, Cesare M. Dalbagno, and Francesco Fumagalli. Eeg synthetic data generation using probabilistic diffusion models, 2023. URL https://arxiv.org/abs/2303.06068.

Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai 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.

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

Hao Zhang, Hongfei Ji, Jian Yu, Jie Li, Lingjing Jin, Lingyu Liu, Zhongfei Bai, and Chen Ye. Subject-independent eeg classification based on a hybrid neural network. *Frontiers in Neuroscience*, 17:1124089, 2023.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Shunan Zhao and Frank Rudzicz. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 992–996. IEEE, 2015.

Wei Zhao, Xiaolu Jiang, Baocan Zhang, Shixiao Xiao, and Sujun Weng. Ctnet: a convolutional transformer network for eeg-based motor imagery classification. *Scientific reports*, 14(1):20237, 2024.