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

Anonymous Author(s)
Affiliation
Address
email

Contents

1 Implementation and experimental details **3**
1.1 Related work 3
1.2 Mixup strategies 3
1.3 Loss functions 3
1.3.1 Classification loss. 3
1.3.2 Supervised contrastive loss. 4
1.3.3 L1 reconstruction loss. 4

2 Generalization performance and analysis **5**

3 Complete ablation studies **6**
3.1 Decoder input ablations 6
3.2 Classifier head ablations 8
3.3 Encoder and decoder ablations 10
3.4 Loss combinations ablations 12
3.5 Mixup strategies ablations 14

4 Complete statistical reporting results **18**
4.1 Methodology 18
4.2 Results 18
4.2.1 Practical evidence assessments 18
4.2.2 Bayesian evidence assessment 18
4.2.3 Wilcoxon signed rank tests and permutation tests 18

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Appendix overview

This appendix provides full experimental details, supporting results, and statistical analyses that complement the main paper:

- **Section 1: Implementation and experimental details**
  Covers architecture choices, training setups, loss function derivations, and augmentation strategies.
- **Section 2: Generalization performance and analysis**
  Presents extended decoding results across all tasks with seen/unseen splits.
- **Section 3: Complete ablation studies**
  Contains exhaustive ablations over decoder input types, classifier heads, architectural variants, mixup strategies, and loss function choices. Each is accompanied by multi-task results.
- **Section 4: Complete statistical reporting results**
  Details our trend-level statistical framework, including effect size estimation, Bayesian comparison, and win-rate matrix construction.

We include additional sections from the main text (e.g., Related Work) that were moved here due to space constraints.

1 Implementation and experimental details

1.1 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 depth-wise–separable convolutions and lightweight design [Lawhern et al., 2018]. Recent models explore transformers [Liao et al., 2025, Song et al., 2022] 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 and 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 and 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.

1.2 Mixup strategies

Algorithm 1 Temporal Masked Mixup

1: Initialize a binary mask $M \in \{0,1\}^{C \times T}$ with all zeros.
2: Flip each 0 in M to 1 with probability $p = 0.01$.
3: **for** each position in M with value 1 **do**
4: Expand to a temporal window of random length (uniform between min and max size).
5: **end for**
6: 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**).

7: Apply the final mask:
- $0 \to x$ (original input)
- $1 \to \hat{x}$ (DDPM output)
- $-1 \to x_{\text{dec}}$ (decoder output)

1.3 Loss functions

1.3.1 Classification loss.

We use either cross-entropy (CE) or mean squared error (MSE).

CE uses softmax over the classifier logits:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{N} y_c \log \hat{y}_c, \quad \hat{y} = \text{Softmax}(\text{FC}(z))$$

MSE treats classification as regression:

$$\mathcal{L}_{\text{MSE}} = \|\hat{y} - y\|_2^2$$

1.3.2 Supervised contrastive loss.

A projection head maps the normalized embedding z to z_{proj}. We then use the SupCon loss:

$$\mathcal{L}_{\text{SupCon}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)}$$

where τ is the temperature, $P(i)$ are positives, and $A(i)$ includes all non-anchor samples.

1.3.3 L1 reconstruction loss.

The decoder learns to reconstruct the DDPM-denoised signal:

$$\mathcal{L}_{\text{L1}} = \|x_{\text{dec}} - \hat{x}\|_1$$

We also compute an auxiliary loss $\|\hat{x} - x\|_1$ to guide DDPM training but exclude it from the final objective.

2 Generalization performance and analysis

Table 1: Final results across tasks and models

Task Paradigm	Configuration	Seen				Unseen			
		Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
SSVEP	`EEGNet`	89.16 ± 0.57	89.16 ± 0.55	89.16 ± 0.57	99.72 ± 0.04	81.09 ± 9.16	81.58 ± 8.21	81.09 ± 9.16	98.75 ± 1.35
	`MultiDiffNet`	85.08 ± 1.53	84.95 ± 1.66	85.08 ± 1.53	99.37 ± 0.05	**84.72 ± 6.04**	**84.44 ± 6.21**	**84.72 ± 6.04**	**98.90 ± 0.76**
	`MultiDiffNet + Temp. Masked Mixup`	86.79 ± 1.75	86.00 ± 0.00	86.00 ± 0.00	99.00 ± 0.00	85.26 ± 6.94	83.00 ± 0.07	83.00 ± 0.07	99.00 ± 0.01
P300	`EEGNet`	88.79 ± 0.68	78.09 ± 1.73	76.03 ± 1.99	89.43 ± 1.06	87.24 ± 2.01	73.72 ± 5.30	71.40 ± 5.69	85.97 ± 4.19
	`MultiDiffNet`	85.24 ± 1.02	63.06 ± 5.33	60.79 ± 4.04	77.54 ± 4.84	76.26 ± 3.61	62.80 ± 3.19	64.95 ± 2.86	71.71 ± 4.68
MI	`EEGNet`	64.09 ± 5.77	63.62 ± 5.79	64.09 ± 5.77	87.09 ± 3.11	43.89 ± 8.33	42.57 ± 8.35	43.89 ± 8.33	71.95 ± 7.51
	`MultiDiffNet`	59.03 ± 6.73	57.12 ± 7.83	59.03 ± 6.73	83.05 ± 2.97	40.25 ± 4.63	38.82 ± 4.72	40.25 ± 4.63	68.06 ± 4.98
Imag. Speech	`EEGNet`	11.26 ± 2.01	11.19 ± 1.98	11.26 ± 2.02	52.02 ± 0.42	10.61 ± 0.93	10.17 ± 0.75	10.61 ± 0.93	50.33 ± 1.25
	`MultiDiffNet`	**15.55 ± 0.62**	9.61 ± 1.23	**15.53 ± 0.62**	**71.24 ± 1.66**	**11.62 ± 1.29**	7.84 ± 1.32	**11.62 ± 1.29**	**57.15 ± 3.18**

Accuracy (mean ± std) reported separately for both seen-subject (intra-subject) and unseen-subject (cross-subject) test splits. Tasks are ranked by task difficulty.

3 Complete ablation studies

3.1 Decoder input ablations

Table 2: Ablation study of decoder input combinations for *MultiDiffNet*

Decoder Input	SSVEP		P300		Motor Imagery		Imagined Speech	
	Seen	Unseen	Seen	Unseen	Seen	Unseen	Seen	Unseen
`x + x_hat + skips`	**85.86 ± 0.38**	83.12 ± 6.90	85.07 ± 1.03	79.15 ± 3.73	53.08 ± 4.68	38.74 ± 6.58	17.07 ± 1.86	12.37 ± 2.79
`x + x_hat`	85.16 ± 0.95	84.40 ± 6.22	85.62 ± 0.43	78.59 ± 1.56	54.56 ± 1.83	40.13 ± 8.25	**18.58 ± 2.13**	10.48 ± 2.87
`x + skips`	85.63 ± 0.29	81.73 ± 7.29	85.75 ± 0.09	77.93 ± 1.43	54.66 ± 2.58	**40.36 ± 7.30**	17.32 ± 1.40	11.74 ± 2.42
`x_hat + skips`	85.70 ± 0.14	83.55 ± 7.93	85.60 ± 0.57	**81.41 ± 0.98**	54.27 ± 2.88	40.31 ± 7.48	17.19 ± 0.47	12.50 ± 1.72
`skips only`	84.03 ± 1.33	82.59 ± 4.69	84.53 ± 0.82	75.52 ± 4.93	53.77 ± 2.80	37.33 ± 6.58	17.70 ± 2.30	11.11 ± 2.24
`z only`	85.08 ± 1.53	**84.72 ± 6.04**	85.35 ± 1.12	79.47 ± 0.54	55.85 ± 2.80	39.24 ± 7.95	15.55 ± 0.62	11.62 ± 1.29
`z + x`	84.89 ± 0.55	82.05 ± 6.57	85.42 ± 0.09	79.07 ± 3.97	**56.89 ± 2.92**	40.02 ± 7.01	18.08 ± 1.74	10.86 ± 2.81
`z + x_hat`	85.16 ± 0.22	82.59 ± 8.17	**85.88 ± 0.38**	79.47 ± 0.64	55.61 ± 3.04	39.64 ± 6.02	17.57 ± 0.37	**12.88 ± 1.07**
`z + skips`	85.31 ± 0.59	83.55 ± 5.60	85.79 ± 0.66	79.09 ± 5.66	52.08 ± 3.04	37.01 ± 5.67	17.83 ± 1.09	10.23 ± 1.93

Accuracy (mean ± std) reported separately for seen- and unseen-subject test splits across four tasks.

Table 3: Ablation study of decoder input combinations for *MultiDiffNet* in the ssvep task

Configuration	SSVEP (Seen)				SSVEP (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`x + x_hat + skips`	**85.86 ± 0.38**	**85.84 ± 0.40**	**85.86 ± 0.38**	**99.50 ± 0.03**	83.12 ± 6.91	82.85 ± 6.92	83.12 ± 6.91	98.79 ± 0.94
`x + x_hat`	85.16 ± 0.95	85.14 ± 0.96	85.16 ± 0.95	99.40 ± 0.06	84.40 ± 6.22	84.14 ± 6.36	84.40 ± 6.22	**98.94 ± 0.80**
`x_hat + skips`	85.70 ± 0.15	85.67 ± 0.14	85.70 ± 0.15	99.45 ± 0.01	83.55 ± 7.93	83.33 ± 8.13	83.55 ± 7.93	98.92 ± 0.84
`x + skips`	85.63 ± 0.29	85.61 ± 0.32	85.63 ± 0.29	99.40 ± 0.03	81.73 ± 7.29	81.52 ± 7.44	81.73 ± 7.29	98.85 ± 0.95
`skips`	84.03 ± 1.33	83.86 ± 1.46	84.03 ± 1.33	99.41 ± 0.02	82.59 ± 4.69	82.22 ± 4.66	82.59 ± 4.69	98.80 ± 0.91
`z only`	85.08 ± 1.53	84.95 ± 1.66	85.08 ± 1.53	99.37 ± 0.05	**84.72 ± 6.04**	**84.44 ± 6.21**	**84.72 ± 6.04**	98.90 ± 0.76
`z + x`	84.89 ± 0.55	84.88 ± 0.63	84.89 ± 0.55	99.45 ± 0.08	82.05 ± 6.57	81.93 ± 6.55	82.05 ± 6.57	98.81 ± 0.87
`z + x_hat`	85.16 ± 0.22	85.12 ± 0.15	85.16 ± 0.22	99.45 ± 0.07	82.59 ± 8.17	82.39 ± 8.37	82.59 ± 8.17	98.69 ± 0.95
`z + skips`	85.31 ± 0.59	85.28 ± 0.64	85.31 ± 0.59	99.48 ± 0.03	83.55 ± 5.60	83.10 ± 5.65	83.55 ± 5.60	98.82 ± 0.82

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks.

Table 4: Ablation study of decoder input combinations for *MultiDiffNet* in the P300 task

Configuration	P300 (Seen)				P300 (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`x + x_hat + skips`	85.07 ± 1.03	**67.21 ± 1.80**	**64.55 ± 1.39**	77.87 ± 2.81	79.15 ± 3.73	63.75 ± 4.99	64.99 ± 6.27	71.06 ± 7.65
`x + x_hat`	85.62 ± 0.43	64.19 ± 2.47	61.38 ± 1.99	77.86 ± 2.52	78.59 ± 1.56	63.56 ± 3.74	64.68 ± 4.86	71.54 ± 5.86
`x_hat + skips`	85.60 ± 0.57	65.95 ± 1.10	63.00 ± 1.28	**79.70 ± 0.77**	**81.41 ± 0.98**	64.91 ± 4.68	64.99 ± 6.23	**72.75 ± 7.16**
`x + skips`	85.75 ± 0.09	65.96 ± 0.60	62.87 ± 0.57	79.66 ± 1.29	77.93 ± 1.43	62.76 ± 1.61	64.23 ± 3.87	70.32 ± 2.97
`skips`	84.53 ± 0.82	65.14 ± 0.95	62.79 ± 1.48	77.33 ± 1.39	75.52 ± 4.93	62.34 ± 4.91	64.58 ± 4.17	69.91 ± 6.11
`z only`	85.35 ± 1.12	64.39 ± 4.95	61.88 ± 4.10	77.14 ± 4.35	79.47 ± 0.54	63.67 ± 1.58	63.95 ± 1.86	70.79 ± 3.27
`z + x`	85.42 ± 0.09	62.70 ± 1.63	60.17 ± 1.34	75.48 ± 2.57	79.07 ± 3.97	63.93 ± 5.15	64.33 ± 4.48	70.53 ± 5.68
`z + x_hat`	**85.88 ± 0.38**	63.44 ± 1.79	60.60 ± 1.37	76.94 ± 3.20	79.47 ± 0.64	**65.01 ± 3.74**	**66.48 ± 5.47**	72.36 ± 5.08
`z + skips`	85.79 ± 0.66	65.02 ± 2.72	62.05 ± 2.19	78.77 ± 2.18	79.09 ± 5.66	64.69 ± 5.87	65.70 ± 5.82	71.68 ± 7.35

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks.

Table 5: Ablation study of decoder input combinations for *MultiDiffNet* in the MI task

Configuration	MI (Seen)				MI (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`x + x_hat + skips`	53.08 ± 4.68	48.79 ± 5.69	53.08 ± 4.68	80.78 ± 4.91	38.74 ± 6.58	37.74 ± 6.87	38.74 ± 6.58	65.41 ± 6.77
`x + x_hat`	54.56 ± 1.83	52.83 ± 1.61	54.56 ± 1.83	83.00 ± 0.71	40.13 ± 8.25	38.69 ± 8.47	40.13 ± 8.25	66.58 ± 6.26
`x_hat + skips`	54.27 ± 2.88	52.66 ± 4.08	54.27 ± 2.88	81.45 ± 1.68	40.31 ± 7.48	39.13 ± 7.91	40.31 ± 7.48	66.60 ± 6.85
`x + skips`	54.66 ± 2.58	52.09 ± 4.13	54.66 ± 2.58	82.00 ± 1.47	**40.36 ± 7.30**	39.25 ± 7.58	**40.36 ± 7.30**	66.33 ± 6.98
`skips`	53.77 ± 2.80	51.12 ± 3.68	53.77 ± 2.80	**83.80 ± 1.37**	37.33 ± 6.58	36.21 ± 7.14	37.33 ± 6.58	63.85 ± 5.14
`z only`	55.85 ± 2.80	54.25 ± 3.15	55.85 ± 2.80	82.72 ± 0.86	39.24 ± 7.95	38.07 ± 7.86	39.24 ± 7.95	**66.70 ± 7.32**
`z + x`	**56.89 ± 2.92**	**54.60 ± 4.39**	**56.89 ± 2.92**	83.41 ± 1.95	40.02 ± 7.01	**39.41 ± 7.09**	40.02 ± 7.01	65.69 ± 6.78
`z + x_hat`	55.61 ± 3.04	53.07 ± 4.23	55.61 ± 3.04	82.65 ± 1.04	39.64 ± 6.02	38.78 ± 5.78	39.64 ± 6.02	65.44 ± 5.60
`z + skips`	52.08 ± 3.04	47.78 ± 3.57	52.08 ± 3.04	82.62 ± 1.66	37.01 ± 5.67	35.56 ± 6.21	37.01 ± 5.67	64.71 ± 6.48

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks.

Ablation study of decoder input combinations for *MultiDiffNet* in the imagined speech task

Configuration	Imag. Speech (Seen)				Imag. Speech (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`x + x_hat + skips`	17.07 ± 1.86	13.35 ± 2.60	17.05 ± 1.86	71.15 ± 3.16	12.37 ± 2.79	9.76 ± 2.17	12.37 ± 2.79	57.98 ± 4.50
`x + x_hat`	**18.58 ± 2.13**	13.39 ± 2.72	**18.56 ± 2.14**	72.58 ± 1.68	10.48 ± 2.87	7.63 ± 0.81	10.48 ± 2.87	56.64 ± 6.04
`x_hat + skips`	17.19 ± 0.47	12.90 ± 1.56	17.17 ± 0.47	**73.16 ± 0.59**	12.50 ± 1.72	9.02 ± 0.18	12.50 ± 1.72	56.86 ± 4.74
`x + skips`	17.32 ± 1.40	13.92 ± 1.93	17.30 ± 1.39	71.93 ± 2.02	11.74 ± 2.42	8.57 ± 1.97	11.74 ± 2.42	56.47 ± 3.87
`skips`	17.70 ± 2.30	12.91 ± 2.52	17.68 ± 2.32	72.61 ± 1.56	11.11 ± 2.24	9.59 ± 2.13	11.11 ± 2.24	56.24 ± 2.03
`z only`	15.55 ± 0.62	9.61 ± 1.23	15.53 ± 0.62	71.24 ± 1.66	11.62 ± 1.29	7.84 ± 1.32	11.62 ± 1.29	57.15 ± 3.18
`z + x`	18.08 ± 1.74	**14.40 ± 1.08**	18.06 ± 1.76	73.08 ± 0.57	10.86 ± 2.81	8.50 ± 1.39	10.86 ± 2.81	55.71 ± 3.62
`z + x_hat`	17.57 ± 0.37	11.16 ± 0.56	17.56 ± 0.37	72.45 ± 1.27	**12.88 ± 1.07**	**8.16 ± 0.48**	**12.88 ± 1.07**	**58.78 ± 1.31**
`z + skips`	17.83 ± 1.09	13.25 ± 1.09	17.80 ± 1.07	72.86 ± 0.78	10.23 ± 1.93	7.72 ± 0.59	10.23 ± 1.93	55.81 ± 4.16

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks.

3.2 Classifier head ablations

Table 6: Ablation study of classifier variants

Classifier Variant	SSVEP		P300		Motor Imagery		Imagined Speech	
	Seen	Unseen	Seen	Unseen	Seen	Unseen	Seen	Unseen
x → EEGNet	**89.16 ± 0.57**	**81.09 ± 9.16**	**88.79 ± 0.68**	**87.24 ± 2.01**	**67.01 ± 5.38**	**46.18 ± 7.21**	11.26 ± 2.01	10.61 ± 0.93
x_hat → EEGNet	10.61 ± 4.96	9.19 ± 7.57	79.18 ± 1.32	79.08 ± 1.49	26.63 ± 0.21	25.84 ± 0.43	9.86 ± 3.27	7.70 ± 1.09
deco_out → EEGNet	80.22 ± 0.99	70.62 ± 6.38	84.29 ± 0.77	83.52 ± 1.07	52.63 ± 3.56	35.16 ± 5.48	**16.94 ± 1.28**	**10.73 ± 1.59**
x → FC	10.72 ± 2.01	11.43 ± 5.31	83.30 ± 0.05	83.33 ± 0.00	47.12 ± 0.92	29.48 ± 2.58	16.94 ± 1.97	8.71 ± 0.93
x_hat → FC	3.73 ± 0.25	4.06 ± 0.54	83.33 ± 0.00	83.33 ± 0.00	25.30 ± 0.42	24.77 ± 0.33	10.62 ± 0.02	8.96 ± 0.94
deco_out → FC	77.62 ± 1.44	69.55 ± 6.77	85.10 ± 0.35	84.11 ± 1.01	50.64 ± 2.08	33.74 ± 4.12	**17.95 ± 1.39**	8.71 ± 1.07
z → FC	**85.08 ± 1.53**	**84.72 ± 6.04**	**85.35 ± 0.35**	**84.12 ± 1.01**	**55.85 ± 2.80**	**39.24 ± 7.95**	15.55 ± 0.62	**11.61 ± 1.29**

Accuracy (mean ± std) reported separately for seen- and unseen-subject test splits across four tasks.

Table 7: Ablation study of `classifier_variant` and `classifier_input` combinations on the SSVEP task

Configuration	SSVEP (Seen)				SSVEP (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
eegn_clsf__x	**89.16 ± 0.57**	**89.16 ± 0.55**	**89.16 ± 0.57**	**99.72 ± 0.04**	81.09 ± 9.16	81.58 ± 8.21	81.09 ± 9.16	98.75 ± 1.35
eegn_clsf__x_hat	10.61 ± 4.96	8.73 ± 4.32	10.61 ± 4.96	59.86 ± 5.14	9.19 ± 7.57	8.15 ± 6.91	9.19 ± 7.57	59.39 ± 8.07
eegn_clsf__deco_out	80.23 ± 0.99	80.19 ± 1.01	80.23 ± 0.99	99.35 ± 0.03	70.62 ± 6.38	69.69 ± 6.48	70.62 ± 6.38	98.21 ± 1.26
fc_clsf__x	10.72 ± 2.01	8.91 ± 1.93	10.72 ± 2.01	73.43 ± 3.35	11.43 ± 5.31	9.63 ± 5.44	11.43 ± 5.31	70.57 ± 5.03
fc_clsf__x_hat	3.73 ± 0.25	0.68 ± 0.30	3.73 ± 0.25	49.74 ± 0.48	4.06 ± 0.54	1.17 ± 0.32	4.06 ± 0.54	49.43 ± 1.77
fc_clsf__deco_out	77.62 ± 1.44	77.49 ± 1.59	77.62 ± 1.44	99.25 ± 0.06	69.55 ± 6.77	69.56 ± 6.81	69.55 ± 6.77	98.03 ± 1.17
fc_clsf__z	85.08 ± 1.53	84.95 ± 1.66	85.08 ± 1.53	99.37 ± 0.05	**84.72 ± 6.04**	**84.44 ± 6.21**	**84.72 ± 6.04**	**98.90 ± 0.76**

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on both seen- and unseen-subject test splits. Results are averaged across four tasks (“eegn” = EEGNet, “clsf” = classifier, “deco” = decoder).

Table 8: Ablation study of `classifier_variant`, and `classifier_input` combinations in the P300 task

Configuration	P300 (Seen)				P300 (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
eegn_clsf__x	**88.79 ± 0.68**	**78.09 ± 1.73**	**76.03 ± 1.99**	**89.43 ± 1.06**	**87.24 ± 2.01**	**73.72 ± 5.30**	**71.40 ± 5.69**	**85.97 ± 4.19**
eegn_clsf__x_hat	79.18 ± 1.32	48.09 ± 2.56	49.70 ± 1.82	46.94 ± 8.58	79.08 ± 1.49	47.72 ± 1.89	49.41 ± 1.50	45.95 ± 6.80
eegn_clsf__deco_out	84.29 ± 0.77	56.01 ± 7.51	56.08 ± 4.34	66.90 ± 12.00	83.52 ± 1.07	54.02 ± 8.11	54.95 ± 5.26	64.43 ± 9.42
fc_clsf__x	83.30 ± 0.05	45.50 ± 0.06	50.00 ± 0.00	50.08 ± 0.23	83.33 ± 0.00	45.45 ± 0.00	50.00 ± 0.00	50.32 ± 0.38
fc_clsf__x_hat	83.33 ± 0.00	45.45 ± 0.00	50.00 ± 0.00	49.85 ± 0.50	83.33 ± 0.00	45.45 ± 0.00	50.00 ± 0.00	50.31 ± 0.30
fc_clsf__deco_out	85.10 ± 0.35	60.43 ± 2.05	58.48 ± 1.48	74.88 ± 2.73	84.12 ± 1.01	56.01 ± 5.84	55.80 ± 3.95	70.24 ± 6.64
fc_clsf__z	85.35 ± 1.12	64.39 ± 4.95	61.88 ± 4.10	77.14 ± 4.35	79.47 ± 0.54	63.67 ± 1.58	63.95 ± 1.86	70.79 ± 3.27

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (“eegn” = EEGNet, “clsf” = classifier, “deco” = decoder).

Table 9: Ablation study of `classifier_variant`, and `classifier_input` combinations in the MI task

Configuration	MI (Seen)				MI (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`eegn_clsf__x`	**67.01** ± **5.38**	**67.01** ± **5.39**	**67.01** ± **5.38**	**88.66** ± **3.28**	**46.18** ± **7.21**	**45.69** ± **7.30**	**46.18** ± **7.21**	**72.75** ± **5.62**
`eegn_clsf__x_hat`	26.64 ± 0.21	20.09 ± 3.08	26.64 ± 0.21	53.43 ± 1.52	25.84 ± 0.43	18.80 ± 4.43	25.84 ± 0.43	50.57 ± 0.75
`eegn_clsf__deco_out`	52.63 ± 3.56	51.30 ± 3.43	52.63 ± 3.56	79.26 ± 2.99	35.16 ± 5.48	31.97 ± 4.78	35.16 ± 5.48	60.71 ± 5.32
`fc_clsf__x`	47.12 ± 0.92	46.83 ± 1.08	47.12 ± 0.92	73.82 ± 0.72	29.48 ± 2.59	26.58 ± 2.96	29.48 ± 2.59	57.16 ± 3.26
`fc_clsf__x_hat`	25.30 ± 0.42	12.15 ± 3.04	25.30 ± 0.42	50.92 ± 1.51	24.77 ± 0.33	11.37 ± 1.93	24.77 ± 0.33	49.21 ± 1.85
`fc_clsf__deco_out`	50.64 ± 2.08	49.99 ± 2.38	50.64 ± 2.08	76.84 ± 2.48	33.74 ± 4.12	30.64 ± 5.43	33.74 ± 4.12	60.94 ± 4.40
`fc_clsf__z`	55.85 ± 2.80	54.25 ± 3.15	55.85 ± 2.80	82.72 ± 0.86	39.24 ± 7.95	38.07 ± 7.86	39.24 ± 7.95	66.70 ± 7.32

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (“eegn” = EEGNet, “clsf” = classifier, “deco” = decoder).

Table 10: Ablation study of `classifier_variant`, and `classifier_input` combinations in the imagened speech task

Configuration	Imag. Speech (Seen)				Imag. Speech (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`eegn_clsf__x`	11.26 ± 2.01	11.19 ± 1.98	11.26 ± 2.02	52.02 ± 0.42	10.61 ± 0.93	**10.17** ± **0.75**	10.61 ± 0.93	50.33 ± 1.25
`eegn_clsf__x_hat`	9.86 ± 3.27	8.82 ± 2.53	9.85 ± 3.27	50.06 ± 1.59	7.70 ± 1.09	6.15 ± 0.85	7.70 ± 1.09	49.10 ± 3.36
`eegn_clsf__deco_out`	**16.94** ± **1.28**	**14.73** ± **0.81**	**16.92** ± **1.26**	**73.33** ± **0.13**	10.73 ± 1.59	6.33 ± 1.77	10.73 ± 1.59	51.50 ± 1.99
`fc_clsf__x`	16.94 ± 1.97	16.68 ± 1.78	16.93 ± 1.98	68.36 ± 0.93	8.71 ± 0.93	6.31 ± 0.76	8.71 ± 0.93	50.99 ± 0.69
`fc_clsf__x_hat`	10.62 ± 0.02	4.48 ± 1.18	10.61 ± 0.00	51.62 ± 1.79	8.96 ± 0.94	3.48 ± 0.57	8.96 ± 0.94	51.20 ± 0.41
`fc_clsf__deco_out`	17.95 ± 1.39	17.02 ± 1.18	17.95 ± 1.39	74.32 ± 0.97	8.71 ± 1.07	4.46 ± 0.35	8.71 ± 1.07	49.30 ± 1.62
`fc_clsf__z`	15.55 ± 0.62	9.61 ± 1.23	15.53 ± 0.62	71.24 ± 1.66	**11.62** ± **1.29**	7.84 ± 1.32	**11.62** ± **1.29**	**57.15** ± **3.18**

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks.

3.3 Encoder and decoder ablations

Table 11: Ablation study of `ddpm_variant`, `encoder_input`, and `decoder_variant` combinations

Configuration	SSVEP		P300		Motor Imagery		Imagained Speech	
	Seen	Unseen	Seen	Unseen	Seen	Unseen	Seen	Unseen
`u_dp-x-u_deco`	85.08 ± 1.53	84.72 ± 6.04	85.35 ± 1.12	79.47 ± 0.54	**55.85 ± 2.80**	39.24 ± 7.95	15.55 ± 0.62	11.62 ± 1.29
`u_dp-x-n_deco`	85.51 ± 0.63	83.55 ± 5.60	**85.71 ± 0.38**	**80.93 ± 1.52**	53.22 ± 4.58	**40.16 ± 6.17**	18.71 ± 1.43	10.98 ± 1.61
`u_dp-x_h-u_deco`	6.53 ± 1.07	7.58 ± 2.35	80.02 ± 4.40	69.43 ± 4.04	25.20 ± 0.86	26.04 ± 0.07	8.85 ± 0.37	9.47 ± 1.35
`u_dp-x_h-n_deco`	6.72 ± 1.35	7.05 ± 1.63	83.11 ± 0.18	66.39 ± 6.24	28.12 ± 0.99	26.04 ± 0.28	8.97 ± 0.16	8.33 ± 2.75
`n_dp-x-u_deco`	90.29 ± 0.29	84.94 ± 8.16	85.36 ± 1.28	78.84 ± 5.44	55.06 ± 5.31	36.49 ± 2.70	**19.22 ± 2.13**	**13.76 ± 1.25**
`n_dp-x-n_deco`	**90.95 ± 0.83**	**85.58 ± 6.15**	85.49 ± 0.66	79.46 ± 0.72	53.22 ± 4.44	38.54 ± 3.87	19.22 ± 0.20	11.36 ± 1.93

Accuracy (mean ± std) reported separately for seen/unseen-subject test splits across tasks ("dp" = ddpm_variant, "x_h" = x_hat, "deco" = decoder_variant,"u" = use,"n" = not use,).

Table 12: Ablation study of `ddpm_variant`, `encoder_input` and `decoder_variant` combinations in the SSVEP task

Configuration	SSVEP (Seen)				SSVEP (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`u_dp__x__u_deco`	85.08 ± 1.53	84.95 ± 1.66	85.08 ± 1.53	99.37 ± 0.05	84.72 ± 6.04	84.44 ± 6.21	84.72 ± 6.04	98.90 ± 0.76
`u_dp__x__n_deco`	85.51 ± 0.63	85.48 ± 0.69	85.51 ± 0.63	99.44 ± 0.02	83.55 ± 5.60	83.25 ± 6.02	83.55 ± 5.60	98.92 ± 0.99
`u_dp__x_h__u_deco`	6.53 ± 1.07	4.89 ± 0.91	6.53 ± 1.07	55.27 ± 1.84	7.59 ± 2.35	7.11 ± 2.35	7.59 ± 2.35	56.22 ± 2.99
`u_dp__x_h__n_deco`	6.72 ± 1.35	4.38 ± 1.43	6.72 ± 1.35	56.19 ± 3.11	7.05 ± 1.63	6.48 ± 1.48	7.05 ± 1.63	54.26 ± 2.37
`n_dp__x__u_deco`	90.29 ± 0.29	90.36 ± 0.28	90.29 ± 0.29	**99.68 ± 0.04**	84.94 ± 8.16	84.53 ± 8.58	84.94 ± 8.16	**99.11 ± 0.88**
`n_dp__x__n_deco`	**90.95 ± 0.83**	**91.02 ± 0.84**	**90.95 ± 0.83**	99.68 ± 0.01	**85.58 ± 6.15**	**85.32 ± 6.37**	**85.58 ± 6.15**	99.07 ± 0.83

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("dp" = ddpm_variant, "x_h" = x_hat, "deco" = decoder_variant,"u" = use,"n" = not use).

Table 13: Ablation study of `ddpm_variant`, `encoder_input` and `decoder_variant` combinations in the P300 task

Configuration	P300 (Seen)				P300 (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`u_dp__x__u_deco`	85.35 ± 1.12	64.39 ± 4.95	61.88 ± 4.10	77.14 ± 4.35	79.47 ± 0.54	63.67 ± 1.58	63.95 ± 1.86	70.79 ± 3.27
`u_dp__x__n_deco`	**85.71 ± 0.38**	67.06 ± 2.38	64.12 ± 2.35	**79.67 ± 1.76**	**80.94 ± 1.53**	64.74 ± 5.62	65.57 ± 6.99	72.64 ± 7.24
`u_dp__x_h__u_deco`	80.02 ± 4.40	48.44 ± 3.74	51.30 ± 1.76	52.24 ± 4.16	69.43 ± 4.04	50.33 ± 1.63	50.64 ± 1.36	50.71 ± 2.41
`u_dp__x_h__n_deco`	83.11 ± 0.18	45.76 ± 0.42	50.02 ± 0.11	48.80 ± 1.49	66.39 ± 6.24	47.84 ± 2.84	48.03 ± 2.67	46.97 ± 3.74
`n_dp__x__u_deco`	85.36 ± 1.28	**68.45 ± 3.02**	**66.09 ± 3.16**	79.59 ± 2.08	78.84 ± 5.44	**65.21 ± 5.06**	**66.28 ± 3.31**	**73.36 ± 6.13**
`n_dp__x__n_deco`	85.50 ± 0.66	62.95 ± 4.25	60.57 ± 3.45	76.39 ± 3.83	79.46 ± 0.72	64.63 ± 2.54	65.53 ± 3.49	72.86 ± 3.49

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("dp" = ddpm_variant, "x_h" = x_hat, "deco" = decoder_variant,"u" = use,"n" = not use).

Table 14: Ablation study of ddpm_variant, encoder_input and decoder_variant combinations in the MI task

Configuration	MI (Seen)				MI (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
u_dp_x_u_deco	**55.85 ± 2.80**	**54.25 ± 3.15**	**55.85 ± 2.80**	**82.72 ± 0.86**	39.24 ± 7.95	**38.07 ± 7.86**	39.24 ± 7.95	**66.70 ± 7.32**
u_dp_x_n_deco	53.22 ± 4.58	51.30 ± 5.58	53.22 ± 4.58	81.97 ± 1.51	**40.16 ± 6.18**	37.50 ± 4.49	**40.16 ± 6.18**	65.69 ± 6.26
u_dp_x_h_u_deco	25.20 ± 0.86	16.31 ± 1.77	25.20 ± 0.86	51.24 ± 0.50	26.04 ± 0.07	22.09 ± 1.31	26.04 ± 0.07	49.78 ± 0.32
u_dp_x_h_n_deco	28.12 ± 0.99	23.52 ± 3.09	28.12 ± 0.99	52.46 ± 0.57	26.04 ± 0.28	19.57 ± 1.42	26.04 ± 0.28	53.04 ± 2.04
n_dp_x_u_deco	55.06 ± 5.31	52.65 ± 8.20	55.06 ± 5.31	81.31 ± 5.04	36.49 ± 2.70	35.02 ± 2.24	36.49 ± 2.70	62.69 ± 2.92
n_dp_x_n_deco	53.22 ± 4.44	50.50 ± 5.52	53.22 ± 4.44	81.79 ± 1.09	38.54 ± 3.87	36.72 ± 3.79	38.54 ± 3.87	64.42 ± 4.25

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("dp" = ddpm_variant, "x_h" = x_hat, "deco" = decoder_variant,"u" = use,"n" = not use).

Table 15: Ablation study of ddpm_variant, encoder_input and decoder_variant combinations in the imagined speech task

Configuration	Imag. Speech (Seen)				Imag. Speech (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
u_dp_x_u_deco	15.55 ± 0.62	9.61 ± 1.23	15.53 ± 0.62	71.24 ± 1.66	11.62 ± 1.29	7.84 ± 1.32	11.62 ± 1.29	57.15 ± 3.18
u_dp_x_n_deco	18.71 ± 1.43	13.92 ± 2.57	18.69 ± 1.46	72.81 ± 0.13	10.98 ± 1.61	8.06 ± 0.55	10.98 ± 1.61	56.99 ± 4.34
u_dp_x_h_use_deco	8.85 ± 0.37	1.72 ± 0.28	8.84 ± 0.36	51.77 ± 1.23	9.47 ± 1.35	7.09 ± 0.63	9.47 ± 1.35	52.19 ± 2.60
u_dp_x_h_n_deco	8.98 ± 0.16	2.22 ± 0.99	8.96 ± 0.18	49.93 ± 0.32	8.33 ± 2.75	6.21 ± 1.83	8.33 ± 2.75	51.19 ± 3.38
n_dp_x_u_deco	**19.22 ± 2.13**	14.48 ± 1.41	**19.20 ± 2.11**	**74.23 ± 1.24**	**13.76 ± 1.25**	**10.59 ± 0.42**	**13.76 ± 1.25**	**57.67 ± 3.22**
n_dp_x_n_deco	19.22 ± 0.20	**14.78 ± 0.77**	19.19 ± 0.18	73.51 ± 1.08	11.36 ± 1.93	7.95 ± 2.65	11.36 ± 1.93	54.54 ± 0.85

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("dp" = ddpm_variant, "x_h" = x_hat, "deco" = decoder_variant,"u" = use,"n" = not use).

3.4 Loss combinations ablations

Table 16: Ablation study of loss in the SSVEP task

Configuration	SSVEP (Seen)				SSVEP (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`CE_a0.5_b0_g0`	84.27 ± 1.34	84.20 ± 1.39	84.27 ± 1.34	99.39 ± 0.12	82.80 ± 5.98	82.45 ± 6.16	82.80 ± 5.98	98.95 ± 0.79
`CE_a0.5_b0_gsched 0.2`	84.27 ± 1.65	84.14 ± 1.72	84.27 ± 1.65	99.39 ± 0.13	81.62 ± 5.61	81.17 ± 5.81	81.62 ± 5.61	98.87 ± 0.81
`CE_a0.5_bsched 0.05_g0`	86.01 ± 0.86	85.99 ± 0.89	86.01 ± 0.86	99.48 ± 0.02	83.12 ± 5.75	82.68 ± 6.03	83.12 ± 5.75	98.89 ± 0.72
`CE_a0.5_bsched 0.05_gsched 0.2`	**86.40 ± 0.86**	**86.36 ± 0.92**	**86.40 ± 0.86**	**99.51 ± 0.04**	83.23 ± 4.69	82.79 ± 5.00	83.23 ± 4.69	**99.03 ± 0.64**
`CE_a1_b0_g0`	83.49 ± 1.72	83.38 ± 1.79	83.49 ± 1.72	99.32 ± 0.06	84.83 ± 4.91	84.56 ± 5.01	84.83 ± 4.91	99.02 ± 0.73
`CE_a1_b0_gsched 0.2`	84.46 ± 0.78	84.33 ± 0.87	84.46 ± 0.78	99.39 ± 0.08	83.44 ± 5.53	82.93 ± 5.88	83.44 ± 5.53	98.89 ± 0.82
`CE_a1_bsched 0.05_g0`	85.16 ± 0.95	85.14 ± 0.96	85.16 ± 0.95	99.40 ± 0.06	84.40 ± 6.22	84.14 ± 6.36	84.40 ± 6.22	98.94 ± 0.80
`CE_a1_bsched 0.05_gsched 0.2`	85.08 ± 1.53	84.95 ± 1.66	85.08 ± 1.53	99.37 ± 0.05	84.72 ± 6.04	84.44 ± 6.21	84.72 ± 6.04	98.90 ± 0.76
`MSE_a0.5_b0_g0`	85.39 ± 0.68	85.18 ± 0.90	85.39 ± 0.68	98.49 ± 0.10	**85.58 ± 8.39**	**85.49 ± 8.47**	**85.58 ± 8.39**	98.19 ± 1.02
`MSE_a0.5_b0_gsched 0.2`	84.97 ± 0.62	84.77 ± 0.61	84.97 ± 0.62	98.20 ± 0.16	84.83 ± 7.03	84.67 ± 7.26	84.83 ± 7.03	97.60 ± 1.65
`MSE_a0.5_bsched 0.05_g0`	84.50 ± 0.81	84.52 ± 0.81	84.50 ± 0.81	97.55 ± 0.45	80.24 ± 5.75	79.88 ± 5.74	80.24 ± 5.75	96.89 ± 1.73
`MSE_a0.5_bsched 0.05_gsched 0.2`	85.08 ± 0.76	85.09 ± 0.75	85.08 ± 0.76	97.30 ± 0.37	81.09 ± 6.84	80.92 ± 6.93	81.09 ± 6.84	96.69 ± 1.10
`MSE_a1_b0_g0`	85.66 ± 0.59	85.58 ± 0.68	85.66 ± 0.59	98.06 ± 0.11	85.26 ± 7.57	85.00 ± 7.62	85.26 ± 7.57	98.12 ± 1.14
`MSE_a1_b0_gsched 0.2`	85.70 ± 0.67	85.66 ± 0.68	85.70 ± 0.67	98.34 ± 0.12	84.62 ± 8.44	84.26 ± 8.72	84.62 ± 8.44	97.70 ± 1.60
`MSE_a1_bsched .05_g0`	86.21 ± 0.15	86.20 ± 0.11	86.21 ± 0.15	98.08 ± 0.27	83.44 ± 6.83	83.18 ± 6.94	83.44 ± 6.83	97.07 ± 2.09
`MSE_a1_bsched 0.05_gsched 0.2`	85.39 ± 0.15	85.37 ± 0.17	85.39 ± 0.15	98.10 ± 0.12	82.80 ± 6.19	82.48 ± 6.34	82.80 ± 6.19	97.45 ± 1.51

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (“sched” = scheduler to).

Table 17: Ablation study of loss in the P300 task

Configuration	P300 (Seen)				P300 (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
`CE_a0.5_b0_g0`	85.15 ± 1.25	61.34 ± 9.17	60.20 ± 6.39	72.13 ± 8.49	77.61 ± 5.30	59.52 ± 3.58	59.38 ± 2.61	66.68 ± 3.94
`CE_a0.5_b0_gsched 0.2`	84.96 ± 1.07	61.36 ± 8.67	60.18 ± 6.04	72.52 ± 8.40	77.15 ± 4.27	59.83 ± 3.08	60.14 ± 2.78	65.42 ± 3.51
`CE_a0.5_bsched 0.05_g0`	85.36 ± 1.01	62.19 ± 7.03	60.36 ± 4.98	76.10 ± 5.36	75.06 ± 2.99	61.63 ± 1.48	64.22 ± 1.48	70.90 ± 2.06
`CE_a0.5_bsched 0.05_gsched 0.2`	85.51 ± 1.09	65.01 ± 3.74	62.21 ± 3.11	78.48 ± 4.32	76.46 ± 1.12	**63.76 ± 0.81**	**66.84 ± 2.65**	**73.06 ± 2.84**
`CE_a1_b0_g0`	84.56 ± 0.95	65.49 ± 2.69	63.26 ± 2.83	77.34 ± 1.00	**80.18 ± 0.50**	63.07 ± 3.07	63.13 ± 4.36	71.95 ± 4.05
`CE_a1_b0_gsched 0.2`	83.91 ± 1.49	65.53 ± 1.16	**63.65 ± 1.92**	75.12 ± 0.99	78.19 ± 3.49	63.21 ± 3.97	63.99 ± 3.60	70.67 ± 5.66
`CE_a1_bsched 0.05_g0`	85.46 ± 0.75	**66.28 ± 2.44**	63.42 ± 2.23	**78.86 ± 2.41**	79.22 ± 2.47	63.35 ± 3.07	63.66 ± 2.95	71.34 ± 4.79
`CE_a1_bsched 0.05_gsched 0.2`	85.35 ± 1.12	64.39 ± 4.95	61.88 ± 4.10	77.14 ± 4.35	79.47 ± 0.54	63.67 ± 1.58	63.95 ± 1.86	70.79 ± 3.27
`MSE_a0.5_b0_g0`	84.70 ± 1.14	58.84 ± 9.64	58.44 ± 6.15	67.96 ± 9.11	67.08 ± 13.35	54.56 ± 7.66	57.93 ± 4.25	60.24 ± 4.85
`MSE_a0.5_b0_gsched 0.2`	83.90 ± 0.40	55.75 ± 7.38	55.99 ± 4.34	67.58 ± 8.14	63.50 ± 9.96	52.50 ± 6.04	56.70 ± 3.72	59.72 ± 4.79
`MSE_a0.5_bsched 0.05_g0`	**85.69 ± 0.65**	63.41 ± 3.27	60.73 ± 2.50	75.94 ± 2.75	74.81 ± 3.74	61.87 ± 1.89	65.58 ± 4.44	69.66 ± 5.34
`MSE_a0.5_bsched 0.05_gsched 0.2`	85.67 ± 0.52	65.01 ± 2.49	62.09 ± 2.09	75.30 ± 5.37	78.17 ± 1.62	63.36 ± 1.79	64.83 ± 3.56	68.73 ± 4.26
`MSE_a1_b0_g0`	84.24 ± 0.69	58.50 ± 9.22	58.31 ± 5.88	68.51 ± 8.11	68.46 ± 10.36	54.56 ± 6.82	56.33 ± 5.40	59.00 ± 6.62
`MSE_a1_b0_gsched 0.2`	84.68 ± 0.29	61.65 ± 1.34	59.45 ± 1.03	71.05 ± 1.04	72.24 ± 2.85	59.71 ± 3.64	63.04 ± 4.17	67.29 ± 5.38
`MSE_a1_bsched 0.05_g0`	85.20 ± 1.15	66.12 ± 1.40	63.29 ± 1.00	75.87 ± 2.28	72.83 ± 2.19	61.17 ± 1.95	65.55 ± 3.36	69.81 ± 3.62
`MSE_a1_bsched 0.05_gsched 0.2`	85.31 ± 0.56	64.17 ± 1.70	61.41 ± 1.32	75.70 ± 3.09	74.35 ± 2.84	61.78 ± 2.66	65.57 ± 5.00	69.59 ± 5.21

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (“sched” = scheduler to).

Table 18: Ablation study of loss in the MI task

Configuration	**MI (Seen)**				**MI (Unseen)**			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
CE_a0.5_b0_g0	51.84 ± 2.36	47.19 ± 3.13	51.84 ± 2.36	81.38 ± 2.42	39.15 ± 7.00	37.98 ± 6.49	39.15 ± 7.00	66.57 ± 6.63
CE_a0.5_b0_gsched 0.2	53.87 ± 2.48	50.13 ± 3.19	53.87 ± 2.48	81.44 ± 1.65	39.99 ± 6.23	39.21 ± 5.71	39.99 ± 6.23	66.25 ± 6.73
CE_a0.5_bsched 0.05_g0	55.36 ± 2.11	53.21 ± 2.62	55.36 ± 2.11	**83.07 ± 1.54**	40.80 ± 6.38	39.21 ± 5.99	40.80 ± 6.38	67.29 ± 6.94
CE_a0.5_bsched 0.05_gsched 0.2	54.76 ± 3.08	52.67 ± 4.43	54.76 ± 3.08	83.06 ± 1.33	40.48 ± 6.61	39.39 ± 6.28	40.48 ± 6.61	**67.68 ± 6.98**
CE_a1_b0_g0	52.38 ± 3.19	48.02 ± 3.37	52.38 ± 3.19	81.50 ± 2.66	39.15 ± 4.69	37.77 ± 4.16	39.15 ± 4.69	65.42 ± 5.19
CE_a1_b0_gsched 0.2	51.44 ± 3.86	46.19 ± 4.98	51.44 ± 3.86	80.63 ± 3.05	39.18 ± 4.31	37.85 ± 3.56	39.18 ± 4.31	65.29 ± 5.10
CE_a1_bsched 0.05_g0	54.46 ± 2.59	52.02 ± 3.63	54.46 ± 2.59	82.65 ± 1.20	40.05 ± 7.45	38.91 ± 7.38	40.05 ± 7.45	67.01 ± 6.95
CE_a1_bsched 0.05_gsched 0.2	55.85 ± 2.80	54.25 ± 3.15	55.85 ± 2.80	82.72 ± 0.86	39.24 ± 7.95	38.07 ± 7.86	39.24 ± 7.95	66.70 ± 7.32
MSE_a0.5_b0_g0	53.92 ± 5.58	51.22 ± 8.61	53.92 ± 5.58	79.86 ± 4.12	37.33 ± 2.98	34.85 ± 1.05	37.33 ± 2.98	63.60 ± 4.09
MSE_a0.5_b0_gsched 0.2	53.62 ± 5.23	50.63 ± 7.83	53.62 ± 5.23	79.80 ± 4.09	37.59 ± 2.99	34.90 ± 1.34	37.59 ± 2.99	63.60 ± 3.91
MSE_a0.5_bsched 0.05_g0	57.74 ± 1.48	55.92 ± 2.94	57.74 ± 1.48	79.98 ± 2.41	40.89 ± 5.00	40.20 ± 4.83	40.89 ± 5.00	63.96 ± 3.68
MSE_a0.5_bsched 0.05_gsched 0.2	**59.67 ± 1.72**	**58.45 ± 2.70**	**59.67 ± 1.72**	81.22 ± 1.05	40.60 ± 4.03	39.56 ± 3.52	40.60 ± 4.03	63.57 ± 3.07
MSE_a1_b0_g0	54.07 ± 5.19	51.81 ± 8.05	54.07 ± 5.19	80.34 ± 3.07	37.30 ± 4.89	36.10 ± 4.59	37.30 ± 4.89	62.84 ± 5.47
MSE_a1_b0_gsched 0.2	54.91 ± 5.02	52.24 ± 7.49	54.91 ± 5.02	81.03 ± 3.28	37.33 ± 4.36	36.21 ± 4.11	37.33 ± 4.36	63.15 ± 4.89
MSE_a1_bsched 0.05_g0	57.79 ± 1.30	57.15 ± 1.89	57.79 ± 1.30	81.44 ± 2.13	39.32 ± 3.39	38.31 ± 2.84	39.32 ± 3.39	63.34 ± 3.57
MSE_a1_bsched 0.05_gsched 0.2	59.23 ± 1.58	57.85 ± 3.33	59.23 ± 1.58	82.16 ± 1.11	**41.44 ± 6.03**	**40.37 ± 6.26**	**41.44 ± 6.03**	65.18 ± 4.76

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("sched" = scheduler to).

Table 19: Ablation study of loss in the imagined speech task

Configuration	**Imag. Speech (Seen)**				**Imag. Speech (Unseen)**			
	Acc	F1	Recall	AUC	Acc	F1	Recall	AUC
CE_a0.5_b0_g0	17.83 ± 1.38	13.40 ± 2.54	17.80 ± 1.35	70.88 ± 1.42	13.26 ± 0.82	9.98 ± 1.14	13.26 ± 0.82	57.90 ± 2.93
CE_a0.5_b0_gsched 0.2	19.09 ± 0.75	**15.08 ± 0.99**	19.07 ± 0.78	72.13 ± 1.37	11.99 ± 2.87	9.08 ± 2.30	11.99 ± 2.87	**59.72 ± 1.83**
CE_a0.5_bsched 0.05_g0	18.46 ± 0.68	15.04 ± 0.64	18.43 ± 0.64	71.98 ± 0.72	11.24 ± 0.99	8.54 ± 1.38	11.24 ± 0.99	56.68 ± 3.58
CE_a0.5_bsched 0.05_gsched 0.2	**19.60 ± 1.62**	13.61 ± 1.07	**19.57 ± 1.59**	**73.49 ± 0.24**	9.34 ± 3.16	5.83 ± 1.34	9.34 ± 3.16	56.33 ± 3.92
CE_a1_b0_g0	18.08 ± 1.28	12.39 ± 0.35	18.06 ± 1.25	72.35 ± 0.89	11.49 ± 1.96	9.18 ± 2.13	11.49 ± 1.96	59.17 ± 1.89
CE_a1_b0_gsched 0.2	17.95 ± 0.70	13.52 ± 1.18	17.93 ± 0.71	73.07 ± 0.40	11.87 ± 1.53	8.18 ± 2.56	11.87 ± 1.53	59.12 ± 2.40
CE_a1_bsched 0.05_g0	17.44 ± 1.70	11.50 ± 0.76	17.42 ± 1.72	71.56 ± 2.03	12.12 ± 2.23	8.78 ± 1.11	12.12 ± 2.23	57.33 ± 3.33
CE_a1_bsched 0.05_gsched 0.2	15.55 ± 0.62	9.61 ± 1.23	15.53 ± 0.62	71.24 ± 1.66	11.62 ± 1.29	7.84 ± 1.32	11.62 ± 1.29	57.15 ± 3.18
MSE_a0.5_b0_g0	16.69 ± 1.32	11.20 ± 0.62	16.67 ± 1.35	66.83 ± 5.71	11.99 ± 0.47	8.54 ± 1.26	11.99 ± 0.47	55.95 ± 1.92
MSE_a0.5_b0_gsched 0.2	16.82 ± 1.98	11.22 ± 0.64	16.79 ± 1.96	64.87 ± 4.18	11.87 ± 1.39	7.68 ± 2.71	11.87 ± 1.39	54.02 ± 2.17
MSE_a0.5_bsched 0.05_g0	17.19 ± 0.81	10.68 ± 1.75	17.17 ± 0.78	70.09 ± 5.01	9.60 ± 1.53	5.44 ± 0.55	9.60 ± 1.53	52.43 ± 4.56
MSE_a0.5_bsched 0.05_gsched 0.2	17.32 ± 0.80	12.48 ± 1.70	17.33 ± 0.81	61.80 ± 7.84	11.62 ± 1.17	8.85 ± 1.36	11.62 ± 1.17	53.32 ± 0.93
MSE_a1_b0_g0	16.43 ± 0.98	12.53 ± 0.65	16.43 ± 0.99	62.13 ± 5.47	**13.51 ± 2.36**	**10.05 ± 1.92**	**13.51 ± 2.36**	53.55 ± 2.59
MSE_a1_b0_gsched 0.2	16.06 ± 2.01	11.90 ± 1.40	16.10 ± 2.04	65.92 ± 4.70	13.26 ± 1.93	9.80 ± 1.77	13.26 ± 1.93	54.53 ± 1.07
MSE_a1_bsched 0.05_g0	17.32 ± 1.11	12.89 ± 1.93	17.31 ± 1.11	61.35 ± 4.49	9.97 ± 2.52	6.64 ± 1.24	9.97 ± 2.52	50.25 ± 1.56
MSE_a1_bsched 0.05_gsched 0.2	17.57 ± 0.67	13.19 ± 1.56	17.58 ± 0.68	61.76 ± 4.09	12.12 ± 2.70	7.66 ± 2.48	12.12 ± 2.70	48.50 ± 1.67

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks ("sched" = scheduler to).

3.5 Mixup strategies ablations

Table 20: Ablation study of mixup in the SSVEP task

Configuration (mixup layer/ warmup epoch/ random ratio)	**SSVEP (Seen)**				**SSVEP (Unseen)**			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
-1 / 100 / No	86.79 ± 1.75	86.75 ± 1.83	86.79 ± 1.75	**99.62 ± 0.02**	**85.26 ± 6.94**	85.02 ± 7.05	**85.26 ± 6.94**	**99.19 ± 0.81**
-1 / 100 / Yes	86.75 ± 0.78	86.74 ± 0.78	86.75 ± 0.78	99.60 ± 0.02	85.15 ± 7.46	**85.09 ± 7.32**	85.15 ± 7.46	99.16 ± 0.92
-1 / 150 / No	86.87 ± 1.77	86.86 ± 1.81	86.87 ± 1.77	99.55 ± 0.07	84.40 ± 7.65	84.07 ± 7.87	84.40 ± 7.65	99.02 ± 1.13
-1 / 150 / Yes	86.60 ± 0.42	86.62 ± 0.44	86.60 ± 0.42	99.54 ± 0.07	84.29 ± 7.09	84.08 ± 7.13	84.29 ± 7.09	99.06 ± 1.09
0 / 100 / No	86.60 ± 1.07	86.59 ± 1.08	86.60 ± 1.07	99.46 ± 0.13	84.51 ± 7.68	84.14 ± 7.92	84.51 ± 7.68	98.99 ± 1.12
0 / 100 / Yes	86.60 ± 1.12	86.56 ± 1.09	86.60 ± 1.12	99.58 ± 0.04	84.72 ± 6.48	84.54 ± 6.51	84.72 ± 6.48	99.15 ± 0.92
0 / 150 / No	87.36 ± 1.34	86.63 ± 0.54	86.64 ± 0.53	99.46 ± 0.11	83.65 ± 8.38	83.46 ± 8.44	83.65 ± 8.38	99.00 ± 1.11
0 / 150 / Yes	**87.84 ± 0.86**	**87.89 ± 0.83**	**87.84 ± 0.86**	99.58 ± 0.01	85.04 ± 7.94	84.85 ± 8.10	85.04 ± 7.94	99.11 ± 1.05
1 / 100 / No	85.08 ± 1.62	85.11 ± 1.62	85.08 ± 1.62	99.34 ± 0.13	84.51 ± 6.85	84.35 ± 6.91	84.51 ± 6.85	99.03 ± 0.97
1 / 100 / Yes	84.65 ± 2.36	84.57 ± 2.55	84.65 ± 2.36	99.40 ± 0.11	82.69 ± 5.94	82.52 ± 5.95	82.69 ± 5.94	98.98 ± 0.93
1 / 150 / No	85.47 ± 0.74	85.43 ± 0.82	85.47 ± 0.74	99.42 ± 0.07	83.33 ± 6.99	83.04 ± 7.16	83.33 ± 6.99	98.95 ± 1.12
1 / 150 / Yes	85.16 ± 0.66	85.09 ± 0.72	85.16 ± 0.66	99.42 ± 0.07	83.55 ± 7.21	83.31 ± 7.45	83.55 ± 7.21	99.00 ± 1.16
2 / 100 / No	85.12 ± 1.55	85.17 ± 1.52	85.12 ± 1.55	99.36 ± 0.12	84.19 ± 6.59	83.99 ± 6.61	84.19 ± 6.59	99.00 ± 0.95
2 / 100 / Yes	85.12 ± 2.06	84.90 ± 2.59	85.12 ± 2.06	99.44 ± 0.15	84.40 ± 7.59	84.26 ± 7.71	84.40 ± 7.59	99.12 ± 0.79
2 / 150 / No	85.47 ± 0.74	85.43 ± 0.82	85.47 ± 0.74	99.42 ± 0.07	83.33 ± 6.99	82.38 ± 6.55	83.33 ± 6.99	98.95 ± 1.12
2 / 150 / Yes	84.89 ± 0.91	84.80 ± 0.97	84.89 ± 0.91	99.43 ± 0.05	83.97 ± 7.70	83.74 ± 7.94	83.97 ± 7.70	98.99 ± 1.15
3 / 100 / No	84.77 ± 2.16	84.72 ± 2.28	84.77 ± 2.16	99.36 ± 0.12	83.01 ± 5.99	82.84 ± 6.01	83.01 ± 5.99	99.03 ± 0.97
3 / 100 / Yes	84.65 ± 2.36	84.57 ± 2.55	84.65 ± 2.36	99.40 ± 0.11	82.69 ± 5.94	82.52 ± 5.95	82.69 ± 5.94	98.98 ± 0.93
3 / 150 / No	85.47 ± 0.74	85.43 ± 0.82	85.47 ± 0.74	99.42 ± 0.07	83.33 ± 6.99	83.04 ± 7.16	83.33 ± 6.99	98.95 ± 1.12
3 / 150 / Yes	85.47 ± 0.74	85.43 ± 0.82	85.47 ± 0.74	99.42 ± 0.07	83.33 ± 6.99	83.04 ± 7.16	83.33 ± 6.99	98.95 ± 1.12
4 / 100 / No	85.70 ± 1.50	85.70 ± 1.61	85.70 ± 1.50	99.42 ± 0.13	83.76 ± 8.14	83.60 ± 8.15	83.76 ± 8.14	98.98 ± 1.07
4 / 100 / Yes	84.85 ± 2.53	84.84 ± 2.78	84.85 ± 2.53	99.39 ± 0.10	82.26 ± 6.60	82.14 ± 6.54	82.26 ± 6.60	99.02 ± 0.86
4 / 150 / No	85.78 ± 1.34	85.81 ± 1.44	85.78 ± 1.34	99.43 ± 0.08	83.23 ± 6.91	83.01 ± 6.78	83.23 ± 6.91	99.02 ± 1.03
4 / 150 / Yes	85.55 ± 0.82	85.56 ± 0.95	85.55 ± 0.82	99.46 ± 0.11	83.33 ± 6.99	82.99 ± 7.24	83.33 ± 6.99	99.08 ± 0.91

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (mixup layer: -1 = temporal mixup at input, 0 = weighted average at input, 1/2/3 = weighted average after first/second/third encoder layer, 4 = weighted average after attention pooling. warmup epoch: number of epochs to train the generators before training the classifier. random ratio: No = equal possibility on choosing ddpm or decoder out for temporal mixup/equal weight for weighted average mixup, Yes = beta/dirichlet distribution with b=0.2 for a random ratio more heavily tilted towards one of the mixup candidates).

Table 21: Ablation study of mixup in the P300 task

Configuration (mixup layer/ warmup epoch/ random ratio)	**P300 (Seen)**				**P300 (Unseen)**			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
-1 / 100 / No	85.61 $\pm$ 0.52	67.30 $\pm$ 1.54	64.42 $\pm$ 2.02	**80.00** $\pm$ **1.39**	**79.56** $\pm$ **4.43**	65.19 $\pm$ 7.51	66.10 $\pm$ 7.91	72.72 $\pm$ 9.74
-1 / 100 / Yes	85.58 $\pm$ 0.47	**67.89** $\pm$ **0.71**	**64.98** $\pm$ **0.71**	79.94 $\pm$ 1.61	79.51 $\pm$ 4.51	**65.22** $\pm$ **7.57**	66.18 $\pm$ 7.97	73.41 $\pm$ 9.22
-1 / 150 / No	85.70 $\pm$ 0.28	66.38 $\pm$ 0.92	63.30 $\pm$ 0.84	78.55 $\pm$ 1.24	78.05 $\pm$ 6.49	64.40 $\pm$ 7.89	65.74 $\pm$ 7.33	71.71 $\pm$ 9.98
-1 / 150 / Yes	**85.78** $\pm$ **0.13**	66.30 $\pm$ 0.51	63.18 $\pm$ 0.56	79.05 $\pm$ 0.35	79.10 $\pm$ 2.04	64.31 $\pm$ 5.87	65.65 $\pm$ 7.85	71.52 $\pm$ 9.12
0 / 100 / No	85.70 $\pm$ 0.44	66.60 $\pm$ 1.84	63.59 $\pm$ 1.88	78.87 $\pm$ 1.90	78.25 $\pm$ 6.85	64.98 $\pm$ 8.43	66.65 $\pm$ 8.39	72.34 $\pm$ 10.36
0 / 100 / Yes	85.54 $\pm$ 0.19	66.38 $\pm$ 1.54	63.43 $\pm$ 1.47	79.63 $\pm$ 1.54	75.11 $\pm$ 5.28	63.41 $\pm$ 6.66	67.46 $\pm$ 8.57	73.52 $\pm$ 10.37
0 / 150 / No	**85.78** $\pm$ **0.41**	67.00 $\pm$ 1.29	63.91 $\pm$ 1.27	78.73 $\pm$ 1.45	78.10 $\pm$ 6.48	64.64 $\pm$ 8.28	66.21 $\pm$ 8.24	72.06 $\pm$ 10.14
0 / 150 / Yes	85.66 $\pm$ 0.25	66.60 $\pm$ 0.89	63.55 $\pm$ 0.81	79.48 $\pm$ 1.97	75.81 $\pm$ 3.96	63.38 $\pm$ 6.33	66.85 $\pm$ 8.63	72.89 $\pm$ 10.04
1 / 100 / No	85.03 $\pm$ 0.33	62.05 $\pm$ 2.81	59.77 $\pm$ 2.19	76.15 $\pm$ 1.21	70.27 $\pm$ 3.61	59.38 $\pm$ 3.95	64.29 $\pm$ 4.65	68.67 $\pm$ 5.30
1 / 100 / Yes	85.48 $\pm$ 0.19	66.53 $\pm$ 1.64	63.66 $\pm$ 1.84	79.16 $\pm$ 1.17	66.82 $\pm$ 3.51	58.12 $\pm$ 1.03	66.45 $\pm$ 5.63	72.31 $\pm$ 8.49
1 / 150 / No	85.02 $\pm$ 0.91	61.30 $\pm$ 5.42	59.35 $\pm$ 4.08	74.17 $\pm$ 5.60	72.35 $\pm$ 3.03	60.51 $\pm$ 3.10	64.56 $\pm$ 3.92	69.57 $\pm$ 5.41
1 / 150 / Yes	85.23 $\pm$ 0.99	64.38 $\pm$ 5.42	61.93 $\pm$ 4.45	77.76 $\pm$ 3.93	71.72 $\pm$ 2.41	60.98 $\pm$ 3.39	66.22 $\pm$ 4.78	71.21 $\pm$ 6.19
2 / 100 / No	84.55 $\pm$ 0.97	57.31 $\pm$ 7.76	56.75 $\pm$ 4.83	71.05 $\pm$ 10.48	70.83 $\pm$ 1.12	58.46 $\pm$ 0.39	62.09 $\pm$ 1.45	66.24 $\pm$ 2.34
2 / 100 / Yes	84.84 $\pm$ 0.80	61.74 $\pm$ 5.07	59.71 $\pm$ 3.79	76.45 $\pm$ 3.50	72.20 $\pm$ 1.37	60.54 $\pm$ 2.63	64.91 $\pm$ 4.27	70.24 $\pm$ 5.46
2 / 150 / No	84.73 $\pm$ 0.75	60.25 $\pm$ 4.67	58.52 $\pm$ 3.42	74.52 $\pm$ 5.66	74.35 $\pm$ 0.65	61.03 $\pm$ 2.28	63.97 $\pm$ 4.40	68.81 $\pm$ 5.90
2 / 150 / Yes	84.87 $\pm$ 0.79	62.09 $\pm$ 5.56	60.04 $\pm$ 4.15	74.71 $\pm$ 5.79	73.87 $\pm$ 0.18	61.51 $\pm$ 2.21	65.20 $\pm$ 4.06	70.76 $\pm$ 5.20
3 / 100 / No	84.86 $\pm$ 0.44	61.05 $\pm$ 1.52	58.93 $\pm$ 1.09	76.08 $\pm$ 0.75	70.95 $\pm$ 4.27	60.24 $\pm$ 4.92	65.37 $\pm$ 5.98	69.56 $\pm$ 7.12
3 / 100 / Yes	85.16 $\pm$ 0.17	65.07 $\pm$ 2.10	62.40 $\pm$ 2.02	78.23 $\pm$ 1.34	71.84 $\pm$ 6.59	62.04 $\pm$ 7.10	**67.89** $\pm$ **7.38**	73.22 $\pm$ 9.40
3 / 150 / No	84.99 $\pm$ 0.41	62.08 $\pm$ 1.94	59.77 $\pm$ 1.62	76.95 $\pm$ 1.70	72.01 $\pm$ 4.05	61.35 $\pm$ 5.41	66.56 $\pm$ 6.89	71.09 $\pm$ 8.62
3 / 150 / Yes	85.27 $\pm$ 0.45	62.82 $\pm$ 3.39	60.39 $\pm$ 2.83	77.27 $\pm$ 1.96	73.99 $\pm$ 4.67	62.18 $\pm$ 5.96	66.33 $\pm$ 7.83	71.19 $\pm$ 10.09
4 / 100 / No	85.47 $\pm$ 0.11	65.16 $\pm$ 2.15	62.30 $\pm$ 1.91	78.26 $\pm$ 1.36	72.84 $\pm$ 5.44	62.07 $\pm$ 5.67	67.37 $\pm$ 7.83	72.55 $\pm$ 8.94
4 / 100 / Yes	85.48 $\pm$ 0.07	65.64 $\pm$ 1.41	62.72 $\pm$ 1.34	79.56 $\pm$ 0.34	76.03 $\pm$ 0.72	63.23 $\pm$ 4.99	67.16 $\pm$ 9.38	**73.91** $\pm$ **9.71**
4 / 150 / No	85.00 $\pm$ 0.91	62.81 $\pm$ 6.22	60.69 $\pm$ 4.74	75.11 $\pm$ 6.29	74.24 $\pm$ 2.88	61.90 $\pm$ 3.60	65.40 $\pm$ 4.55	71.28 $\pm$ 5.96
4 / 150 / Yes	85.26 $\pm$ 0.59	64.19 $\pm$ 3.92	61.60 $\pm$ 3.19	77.58 $\pm$ 2.92	75.38 $\pm$ 4.29	62.57 $\pm$ 4.95	65.28 $\pm$ 4.82	71.59 $\pm$ 6.93

Mean$\pm$std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (mixup layer: -1 = temporal mixup at input, 0 = weighted average at input, 1/2/3 = weighted average after first/second/third encoder layer, 4 = weighted average after attention pooling. warmup epoch: number of epochs to train the generators before training the classifier. random ratio: No = equal possibility on choosing ddpm or decoder out for temporal mixup/equal weight for weighted average mixup, Yes = beta/dirichlet distribution with b=0.2 for a random ratio more heavily tilted towards one of the mixup candidates).

Table 22: Ablation study of mixup in the MI task

Configuration (mixup layer/ warmup epoch/ random ratio)	MI (Seen)				MI (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
-1 / 100 / No	57.69 ± 3.27	55.70 ± 4.59	57.69 ± 3.27	84.97 ± 1.19	36.78 ± 5.22	35.94 ± 5.62	36.78 ± 5.22	63.44 ± 5.17
-1 / 100 / Yes	58.33 ± 2.10	56.58 ± 3.53	58.33 ± 2.10	84.52 ± 1.20	36.23 ± 5.56	35.56 ± 5.90	36.23 ± 5.56	63.44 ± 5.01
-1 / 150 / No	58.13 ± 4.14	56.10 ± 5.03	58.13 ± 4.14	85.32 ± 1.64	35.53 ± 3.03	34.51 ± 3.21	35.53 ± 3.03	62.92 ± 3.83
-1 / 150 / Yes	57.44 ± 4.36	55.81 ± 5.18	57.44 ± 4.36	85.03 ± 2.00	36.49 ± 5.74	35.68 ± 6.36	36.49 ± 5.74	63.62 ± 5.23
0 / 100 / No	62.75 ± 6.07	62.71 ± 5.90	62.75 ± 6.07	85.84 ± 4.03	36.83 ± 8.41	36.01 ± 8.52	36.83 ± 8.41	63.51 ± 6.71
0 / 100 / Yes	62.25 ± 3.33	62.32 ± 3.18	62.25 ± 3.33	86.28 ± 2.31	38.14 ± 8.56	37.99 ± 8.69	38.14 ± 8.56	63.64 ± 5.84
0 / 150 / No	**63.44 ± 3.44**	**63.44 ± 3.36**	**63.44 ± 3.44**	**87.20 ± 2.69**	35.45 ± 6.82	34.63 ± 6.56	35.45 ± 6.82	62.40 ± 5.07
0 / 150 / Yes	62.95 ± 5.94	62.97 ± 5.95	62.95 ± 5.94	86.71 ± 3.13	36.17 ± 6.50	35.66 ± 6.94	36.17 ± 6.50	63.10 ± 5.78
1 / 100 / No	59.38 ± 3.80	59.23 ± 3.95	59.38 ± 3.80	83.61 ± 2.36	36.08 ± 4.55	35.17 ± 5.20	36.08 ± 4.55	63.93 ± 5.77
1 / 100 / Yes	61.56 ± 3.06	61.43 ± 3.04	61.56 ± 3.06	85.52 ± 1.81	37.15 ± 7.61	37.02 ± 7.84	37.15 ± 7.61	63.38 ± 6.53
1 / 150 / No	60.37 ± 2.25	60.50 ± 2.29	60.37 ± 2.25	84.62 ± 1.70	36.34 ± 6.36	35.71 ± 6.82	36.34 ± 6.36	62.81 ± 4.85
1 / 150 / Yes	62.35 ± 4.00	62.42 ± 4.01	62.35 ± 4.00	86.41 ± 2.23	36.63 ± 5.51	36.44 ± 5.79	36.63 ± 5.51	63.27 ± 5.17
2 / 100 / No	59.13 ± 1.21	58.48 ± 1.39	59.13 ± 1.21	84.00 ± 1.30	36.95 ± 5.34	36.40 ± 5.15	36.95 ± 5.34	63.50 ± 5.68
2 / 100 / Yes	58.58 ± 2.26	58.41 ± 2.03	58.58 ± 2.26	83.91 ± 2.18	37.82 ± 5.87	37.69 ± 5.88	37.82 ± 5.87	63.69 ± 5.22
2 / 150 / No	57.74 ± 1.79	57.12 ± 1.01	57.74 ± 1.79	83.61 ± 1.38	**38.83 ± 7.09**	**38.71 ± 7.33**	**38.83 ± 7.09**	**64.79 ± 6.40**
2 / 150 / Yes	61.16 ± 2.46	61.03 ± 2.62	61.16 ± 2.46	85.02 ± 0.98	36.46 ± 6.26	36.53 ± 6.41	36.46 ± 6.26	62.82 ± 5.49
3 / 100 / No	59.77 ± 3.53	59.26 ± 3.93	59.77 ± 3.53	83.27 ± 1.17	36.86 ± 6.21	36.42 ± 6.34	36.86 ± 6.21	63.59 ± 6.18
3 / 100 / Yes	60.07 ± 1.21	59.72 ± 0.76	60.07 ± 1.21	84.79 ± 1.44	37.88 ± 7.69	37.73 ± 7.67	37.88 ± 7.69	63.69 ± 5.73
3 / 150 / No	59.67 ± 4.03	59.30 ± 4.29	59.67 ± 4.03	84.50 ± 2.93	36.72 ± 3.88	36.70 ± 4.28	36.72 ± 3.88	63.31 ± 3.79
3 / 150 / Yes	60.81 ± 3.74	60.61 ± 3.77	60.81 ± 3.74	84.25 ± 2.34	37.36 ± 5.16	37.32 ± 5.27	37.36 ± 5.16	63.37 ± 5.07
4 / 100 / No	57.94 ± 2.28	56.67 ± 0.50	57.94 ± 2.28	84.04 ± 1.48	36.40 ± 3.79	36.01 ± 3.93	36.40 ± 3.79	63.75 ± 3.65
4 / 100 / Yes	58.23 ± 4.14	57.97 ± 3.77	58.23 ± 4.14	84.10 ± 2.88	37.15 ± 5.17	36.69 ± 5.45	37.15 ± 5.17	63.87 ± 4.77
4 / 150 / No	58.33 ± 3.37	57.92 ± 2.86	58.33 ± 3.37	84.05 ± 2.06	37.96 ± 5.57	37.06 ± 5.58	37.96 ± 5.57	64.10 ± 4.85
4 / 150 / Yes	57.79 ± 4.99	56.75 ± 4.50	57.79 ± 4.99	84.08 ± 3.49	36.69 ± 6.35	36.61 ± 6.49	36.69 ± 6.35	63.53 ± 5.17

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (mixup layer: -1 = temporal mixup at input, 0 = weighted average at input, 1/2/3 = weighted average after first/second/third encoder layer, 4 = weighted average after attention pooling. warmup epoch: number of epochs to train the generators before training the classifier. random ratio: No = equal possibility on choosing ddpm or decoder out for temporal mixup/equal weight for weighted average mixup, Yes = beta/dirichlet distribution with b=0.2 for a random ratio more heavily tilted towards one of the mixup candidates).

Table 23: Ablation study of mixup in the Imagined Speech task

Configuration (mixup layer/ warmup epoch/ random ratio)	Imagined Speech (Seen)				Imagined Speech (Unseen)			
	Acc (%)	F1 (%)	Recall (%)	AUC (%)	Acc (%)	F1 (%)	Recall (%)	AUC (%)
-1 / 100 / No	17.57 ± 1.17	12.24 ± 3.26	17.55 ± 1.16	73.01 ± 0.91	**12.12 ± 0.38**	7.77 ± 1.27	**12.12 ± 0.38**	55.73 ± 2.77
-1 / 100 / Yes	17.32 ± 2.69	11.18 ± 4.16	17.30 ± 2.69	**73.64 ± 1.56**	10.86 ± 1.43	6.80 ± 0.49	10.86 ± 1.43	**57.63 ± 0.34**
-1 / 150 / No	18.08 ± 0.42	11.48 ± 2.44	18.06 ± 0.44	73.07 ± 0.81	11.87 ± 0.79	7.43 ± 0.65	11.87 ± 0.79	56.82 ± 1.54
-1 / 150 / Yes	18.46 ± 1.07	11.88 ± 3.10	18.43 ± 1.09	73.20 ± 0.72	11.62 ± 1.22	7.19 ± 0.74	11.62 ± 1.22	56.62 ± 1.89
0 / 100 / No	17.19 ± 0.62	11.74 ± 1.58	17.17 ± 0.58	73.14 ± 0.75	11.36 ± 1.00	7.50 ± 0.70	11.36 ± 1.00	56.25 ± 2.15
0 / 100 / Yes	**19.47 ± 0.95**	**14.70 ± 0.60**	**19.44 ± 0.95**	71.90 ± 1.82	11.62 ± 1.43	**9.29 ± 0.89**	11.62 ± 1.43	53.80 ± 2.29
0 / 150 / No	17.57 ± 0.46	10.97 ± 1.65	17.55 ± 0.44	72.96 ± 0.91	11.87 ± 0.79	7.47 ± 0.66	11.87 ± 0.79	56.72 ± 1.71
0 / 150 / Yes	16.31 ± 1.40	11.45 ± 0.58	16.29 ± 1.37	72.38 ± 1.63	10.35 ± 2.09	6.11 ± 1.09	10.35 ± 2.09	54.91 ± 2.51
1 / 100 / No	18.08 ± 1.77	13.16 ± 3.38	18.06 ± 1.79	72.47 ± 3.16	11.24 ± 2.52	8.33 ± 2.28	11.24 ± 2.52	55.49 ± 1.96
1 / 100 / Yes	18.08 ± 0.42	10.55 ± 1.10	18.06 ± 0.44	73.19 ± 0.73	10.73 ± 2.74	7.45 ± 0.65	10.73 ± 2.74	55.25 ± 4.22
1 / 150 / No	17.95 ± 1.71	13.77 ± 3.24	17.93 ± 1.71	72.17 ± 1.62	9.72 ± 2.66	6.59 ± 0.17	9.72 ± 2.66	54.45 ± 3.34
1 / 150 / Yes	17.19 ± 1.11	10.51 ± 1.05	17.17 ± 1.09	73.00 ± 0.88	11.62 ± 1.22	6.55 ± 1.61	11.62 ± 1.22	55.87 ± 3.15
2 / 100 / No	17.45 ± 0.41	12.91 ± 1.93	17.42 ± 0.38	71.57 ± 2.51	10.23 ± 1.97	6.84 ± 0.55	10.23 ± 1.97	55.16 ± 2.83
2 / 100 / Yes	16.94 ± 0.98	11.57 ± 0.78	16.92 ± 0.95	72.98 ± 0.88	11.87 ± 1.43	7.37 ± 2.88	11.87 ± 1.43	55.90 ± 3.18
2 / 150 / No	16.94 ± 0.98	11.96 ± 1.46	16.92 ± 0.95	72.48 ± 1.47	11.11 ± 1.22	5.91 ± 0.97	11.11 ± 1.22	54.66 ± 2.50
2 / 150 / Yes	17.19 ± 1.11	10.51 ± 1.05	17.17 ± 1.09	73.00 ± 0.88	11.62 ± 1.22	6.55 ± 1.61	11.62 ± 1.22	55.87 ± 3.15
3 / 100 / No	17.45 ± 0.70	14.12 ± 0.68	17.42 ± 0.66	71.25 ± 1.86	9.47 ± 1.65	8.07 ± 1.49	9.47 ± 1.65	52.88 ± 1.35
3 / 100 / Yes	17.07 ± 1.03	11.79 ± 1.17	17.05 ± 1.00	73.38 ± 0.86	11.49 ± 1.16	6.76 ± 1.91	11.49 ± 1.16	55.63 ± 2.91
3 / 150 / No	16.06 ± 0.20	11.54 ± 1.04	16.04 ± 0.22	71.23 ± 1.65	11.11 ± 0.79	6.73 ± 1.63	11.11 ± 0.79	53.75 ± 1.70
3 / 150 / Yes	16.44 ± 1.26	9.43 ± 2.92	16.41 ± 1.22	72.50 ± 1.45	11.74 ± 1.31	5.30 ± 1.29	11.74 ± 1.31	57.20 ± 4.91
4 / 100 / No	16.56 ± 1.13	10.92 ± 0.36	16.54 ± 1.09	73.15 ± 0.81	11.49 ± 1.16	6.92 ± 2.16	11.49 ± 1.16	54.95 ± 2.52
4 / 100 / Yes	17.19 ± 1.11	10.51 ± 1.05	17.17 ± 1.09	73.00 ± 0.88	11.62 ± 1.22	6.55 ± 1.61	11.62 ± 1.22	55.87 ± 3.15
4 / 150 / No	16.44 ± 1.26	11.07 ± 0.13	16.41 ± 1.22	71.59 ± 2.89	10.35 ± 2.09	5.38 ± 1.20	10.35 ± 2.09	54.50 ± 2.53
4 / 150 / Yes	17.19 ± 1.11	10.51 ± 1.05	17.17 ± 1.09	73.00 ± 0.88	11.62 ± 1.22	6.55 ± 1.61	11.62 ± 1.22	55.87 ± 3.15

Mean±std is reported separately for Accuracy, F1, Recall, and AUC on seen- and unseen-subject test splits across four tasks (mixup layer: -1 = temporal mixup at input, 0 = weighted average at input, 1/2/3 = weighted average after first/second/third encoder layer, 4 = weighted average after attention pooling. warmup epoch: number of epochs to train the generators before training the classifier. random ratio: No = equal possibility on choosing ddpm or decoder out for temporal mixup/equal weight for weighted average mixup, Yes = beta/dirichlet distribution with b=0.2 for a random ratio more heavily tilted towards one of the mixup candidates).

4 Complete statistical reporting results

4.1 Methodology

EEG decoding suffers from low statistical power due to extreme inter-subject variability, limited trial counts, and costly LOSO evaluation Kukhilava et al. [2025], Huang et al. [2023]. In our runs with three independent seeds, standard tests (Wilcoxon, permutation) often returned p-values near 1.0, masking reproducible gains, which is an expected outcome in such low-n, high-variance settings Vialatte and Cichocki [2008], Nakagawa [2004]. We therefore propose a complementary framework emphasizing effect size estimation and evidence synthesis over binary significance. This approach quantifies improvement magnitude and consistency across seeds and datasets, retaining sensitivity to systematic trends even when classical tests fail, in line with best practices for robust neural decoding Vialatte and Cichocki [2008], Nakagawa [2004].

For each comparison between configurations c_1 and c_2, we compute Cohen's d and its 95% confidence interval:

$$d = \frac{\bar{x}_{c_1} - \bar{x}_{c_2}}{s_{\text{pooled}}}, \quad \text{CI}_{95\%} = d \pm t_{\alpha/2,df} \cdot \text{SE}_d,$$

where the standard error is

$$\text{SE}_d = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}}.$$

A win is established through hierarchical evidence assessment based on cross-seed consistency and effect magnitude. For configurations with complete cross-seed agreement (i.e., all seeds exhibit consistent directional differences), evidence strength is defined as follows: (1) strong evidence requires both a large effect size ($|d| \geq 0.5$) and a meaningful relative improvement of at least 2%; (2) moderate evidence requires either a large effect size ($|d| \geq 0.5$) or a relative improvement of at least 2%; (3) weak evidence requires a medium effect size ($|d| \geq 0.3$); (4) minimal evidence requires a small effect size ($|d| \geq 0.2$). For configurations exhibiting majority cross-seed agreement (i.e., at least 2 out of 3 seeds show consistent direction), all evidence categories are downgraded by one level. Configuration c_1 is considered to exhibit superior performance over c_2 if any evidence category is satisfied.

To summarize model comparisons, we construct a win-loss matrix $\mathbf{W}$ where $W_{ij} = 1$ if configuration i shows evidence of superiority over configuration j. The win rate of configuration c_i is computed as

$$\text{WinRate}_i = \frac{\sum_j W_{ij}}{\sum_j (W_{ij} + W_{ji})}.$$

This matrix supports a global ranking across all configurations.

As a complementary analysis, we compute posterior probabilities $P(\text{left})$, $P(\text{rope})$, and $P(\text{right})$ using the `baycomp` framework with ROPE threshold $\rho = 0.01$. Configuration c_1 is considered to have Bayesian evidence of superiority if $P(\text{right}) > 0.85$.

4.2 Results

4.2.1 Practical evidence assessments

4.2.2 Bayesian evidence assessment

Due to page limitations, only SSVEP task results are shown.

4.2.3 Wilcoxon signed rank tests and permutation tests

Due to minimal statistical significance after correction, only one complete analytical framework applied to SSVEP ablation results is presented as a demonstrative example, constrained by page limitations

Table 24: Practical evidence assessment of decoder input ablations in the SSVEP task

Config	Seen							Unseen						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
x_hat + skips	**6**	0	6	**1.00**	skips x + x_hat z + skips z + x z + x_hat z only	—	Weak:3 Minimal:2 Moderate:1	4	2	6	0.667	x + skips x + x_hat + skips z + x z + x_hat	x + x_hat z only	Minimal:2 Strong:1 Weak:1
x + x_hat + skips	**7**	0	7	**1.00**	skips x + skips x + x_hat z + skips z + x z + x_hat z only	—	Weak:3 Moderate:2 Strong:1 Minimal:1	2	3	5	0.40	x + skips z + x	x + x_hat x_hat + skips z only	Moderate:2
x + skips	5	1	6	0.833	skips x + x_hat z + skips z + x z + x_hat	x + x_hat + skips	Minimal:2 Weak:2 Moderate:1	0	6	6	0.00	—	x + x_hat x + x_hat + skips x_hat + skips z + skips z + x_hat z only	—
z + skips	2	3	5	0.40	skips z + x	x + skips x + x_hat + skips x_hat + skips	Weak:1 Minimal:1	3	2	5	0.60	skips x + skips z + x	x + x_hat z only	Moderate:2 Weak:1
x + x_hat	2	3	5	0.40	skips z + x	x + skips x + x_hat + skips x_hat + skips	Moderate:1 Weak:1	**7**	0	7	**1.00**	skips x + skips x + x_hat + skips x_hat + skips z + skips z + x z + x_hat	—	Moderate:3 Strong:2 Minimal:1 Weak:1
z only	1	2	3	0.333	skips	x + x_hat + skips x_hat + skips	Moderate:1	**7**	0	7	**1.00**	skips x + skips x + x_hat + skips x_hat + skips z + skips z + x z + x_hat	—	Strong:3 Weak:2 Minimal:1 Moderate:1
z + x_hat	1	3	4	0.25	skips	x + skips x + x_hat + skips x_hat + skips	Weak:1	1	3	4	0.25	x + skips	x + x_hat x_hat + skips z only	Minimal:1
z + x	1	5	6	0.167	skips	x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips	Weak:1	0	5	5	0.00	—	x + x_hat x + x_hat + skips x_hat + skips z + skips z only	—
skips	0	8	8	0.00	—	x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x z + x_hat z only	—	0	3	3	0.00	—	x + x_hat z + skips z only	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 25: Practical evidence assessment of classifier type ablations in the SSVEP task

Config	Seen							Unseen						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
eegnet_classifier__x	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:6	**5**	1	6	**0.83**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	fc_classifier__z	Strong:5
fc_classifier__z	**5**	1	6	**0.83**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x	Strong:5	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	—	Strong:6
eegnet_classifier__decoder_out	4	2	6	0.67	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x fc_classifier__z	Strong:4	4	2	6	0.67	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x fc_classifier__z	Strong:3 Weak:1
fc_classifier__decoder_out	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__z	Strong:3	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__z	Strong:3
eegnet_classifier__x_hat	1	4	5	0.20	fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:1	1	5	6	0.17	fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__z	Moderate:1
fc_classifier__x	1	4	5	0.20	fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:1	2	4	6	0.33	eegnet_classifier__x_hat fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Moderate:1 Strong:1
fc_classifier__x_hat	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__z	—	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__z	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 26: Practical evidence assessment of architecture ablations in the SSVEP task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
no_ddpm__x__no_decoder	**5**	0	5	**1.00**	no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:4 Minimal:1	**4**	0	4	**1.00**	use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:3 Moderate:1
no_ddpm__x__use_decoder	4	1	5	0.80	use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder	Strong:4	3	0	3	1.00	use_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:2 Minimal:1
use_ddpm__x__no_decoder	3	2	5	0.60	use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder	Strong:2 Minimal:1	2	3	5	0.40	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__use_decoder	Strong:2
use_ddpm__x__use_decoder	2	3	5	0.40	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder	Strong:2	3	1	4	0.75	use_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder	Strong:2 Moderate:1
use_ddpm__x_hat__no_decoder	1	4	5	0.20	use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Weak:1	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__use_decoder	—
use_ddpm__x_hat__use_decoder	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder	—	1	4	5	0.20	use_ddpm__x_hat__no_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Weak:1

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 27: Practical evidence assessment of mixup ablations in the SSVEP task

Config	**Seen**							**Unseen**						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
layer-1_randNo	**8**	0	8	**1.00**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Strong:5, Moderate:2, Weak:1	**8**	0	8	**1.00**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Strong:5, Moderate:2, Weak:1
layer-1_randYes	**9**	0	9	**1.00**	layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Moderate:3, Strong:3, Weak:2, Minimal:1	**9**	0	9	**1.00**	layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Moderate:3, Strong:3, Weak:2, Minimal:1
layer0_randYes	**8**	0	8	**1.00**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Moderate:4, Weak:2, Minimal:2	**8**	0	8	**1.00**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Moderate:4, Weak:2, Minimal:2
layer0_randNo	**8**	1	9	**0.89**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	layer-1_randYes	Moderate:3, Strong:3, Weak:2	**8**	1	9	**0.89**	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	layer-1_randYes	Moderate:3, Strong:3, Weak:2
layer4_randNo	7	4	11	0.64	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes	Weak:7	7	4	11	0.64	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes	Weak:7
layer2_randYes	3	5	8	0.38	layer1_randYes, layer3_randNo, layer3_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	Weak:3	3	5	8	0.38	layer1_randYes, layer3_randNo, layer3_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	Weak:3
layer1_randYes	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—
layer1_randNo	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—
layer2_randNo	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—
layer3_randNo	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—
layer3_randYes	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—	0	6	6	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo	—
layer4_randYes	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—	0	5	5	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randNo	—

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 28: Practical evidence assessment of loss ablations in the SSVEP task (Part 1 of 2)

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
CE_a0.5_bscheduler to 0.05_gscheduler to 0.2	**14**	0	14	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Strong:5, Moderate:4, Weak:4, Minimal:1	3	6	9	0.33	CE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a1_b0_g0, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, Strong:1	
MSE_a1_bscheduler to 0.05_g0	**13**	0	13	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Moderate:6, Weak:4, Strong:3	4	7	11	0.36	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a1_b0_g0, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,Minimal:1,Strong:1	
CE_a0.5_bscheduler to 0.05_g0	**13**	1	14	**0.93**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_gscheduler to 0.2	Weak:5, Moderate:4, Strong:3, Minimal:1	5	8	13	0.38	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2	Strong:2, Minimal:1, Moderate:1, Weak:1
MSE_a1_b0_g0	11	3	14	0.79	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_g0	Moderate:4, Weak:4, Minimal:2, Strong:1	**12**	0	12	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Strong:6, Moderate:3, Weak:2, Minimal:1
MSE_a1_b0_gscheduler to 0.2	9	3	12	0.75	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_g0	Moderate:4, Weak:2, Minimal:2, Strong:1	8	2	10	0.80	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a1_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	MSE_a0.5_b0_g0, MSE_a1_b0_g0	Moderate:3, Minimal:2, Strong:2, Weak:1
MSE_a0.5_b0_g0	6	4	10	0.60	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_bscheduler to 0.05_g0	Weak:4, Moderate:2	**11**	0	11	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Moderate:7, Strong:2, Minimal:1, Weak:1
MSE_a1_bscheduler to 0.05_gscheduler to 0.2	7	5	12	0.58	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0	Weak:4, Minimal:2, Moderate:1	3	9	12	0.25	CE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2	Strong:2, Moderate:1
CE_a1_bscheduler to 0.05_gscheduler to 0.2	5	4	9	0.56	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_bscheduler to 0.05_g0	Moderate:2, Weak:2, Minimal:1	**9**	0	9	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Strong:5, Weak:2, Moderate:1, Minimal:1
CE_a1_bscheduler to 0.05_g0	5	5	10	0.50	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0	Moderate:4, Minimal:1	9	3	12	0.75	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0	Moderate:3, Strong:3, Weak:3

Table 29: Practical evidence assessment of loss ablations in the SSVEP task (Part 2 of 2)

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	5	7	12	0.42	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Weak:3, Minimal:2	1	13	14	0.07	MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_b0_g0, CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Minimal:1
MSE_a0.5_b0_gscheduler to 0.2	4	6	10	0.40	CE_a0.5_b0_g0, CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2, MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Minimal:2, Weak:2	10	1	11	0.91	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	MSE_a1_b0_g0	Strong:5, Moderate:2, Minimal:2, Weak:1
CE_a0.5_b0_gscheduler to 0.2	1	10	11	0.09	CE_a1_b0_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Weak:1	1	13	14	0.07	MSE_a0.5_bscheduler to 0.05_g0	CE_a0.5_b0_g0, CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Moderate:1
CE_a0.5_b0_g0	1	11	12	0.08	CE_a1_b0_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Moderate:1	3	10	13	0.23	CE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2	CE_a0.5_bscheduler to 0.05_g0,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0	Moderate:2, Strong:1
CE_a1_b0_gscheduler to 0.2	1	11	12	0.08	CE_a1_b0_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Weak:1	6	7	13	0.46	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	CE_a1_b0_g0, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2	Moderate:2, Strong:2, Minimal:1, Weak:1
MSE_a0.5_bscheduler to 0.05_g0	1	11	12	0.08	CE_a1_b0_g0	CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	Weak:1	0	15	15	0.00	—	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2,CE_a1_b0_g0, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—
CE_a1_b0_g0	0	15	15	0.00	—	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2,CE_a1_bscheduler to 0.05_g0,CE_a1_bscheduler to 0.05_gscheduler to 0.2,MSE_a0.5_b0_g0, MSE_a0.5_b0_gscheduler to 0.2,MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_b0_g0, MSE_a1_b0_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	**9**	0	9	**1.00**	CE_a0.5_b0_g0, CE_a0.5_b0_gscheduler to 0.2,CE_a0.5_bscheduler to 0.05_g0, CE_a0.5_bscheduler to 0.05_gscheduler to 0.2, CE_a1_b0_gscheduler to 0.2, MSE_a0.5_bscheduler to 0.05_g0, MSE_a0.5_bscheduler to 0.05_gscheduler to 0.2,MSE_a1_bscheduler to 0.05_g0, MSE_a1_bscheduler to 0.05_gscheduler to 0.2	—	Strong:6, Moderate:2, Weak:1

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 30: Practical evidence assessment of decoder input ablations in the P300 task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
x + skips	**4**	0	4	**1.00**	skips x + x_hat + skips z + x z only	—	Moderate:2 Weak:1 Minimal:1	1	3	4	0.25	skips	x_hat + skips z + x_hat z only	Weak:1
z + x_hat	**5**	0	5	**1.00**	skips x + x_hat x + x_hat + skips z + x z only	—	Minimal:3 Moderate:2	**3**	1	4	**0.75**	skips x + skips x + x_hat	x_hat + skips	Weak:2 Moderate:1
x_hat + skips	2	0	2	1.00	skips x + x_hat + skips	—	Weak:1 Moderate:1	**8**	0	8	**1.00**	skips x + skips x + x_hat x + x_hat + skips z + skips z + x z + x_hat z only	—	Strong:5 Moderate:2 Weak:1
z + skips	3	0	3	1.00	skips x + x_hat + skips z + x	—	Minimal:2 Moderate:1	1	1	2	0.50	skips	x_hat + skips	Strong:1
x + x_hat	3	1	4	0.75	skips x + x_hat + skips z + x	z + x_hat	Minimal:2 Moderate:1	1	3	4	0.25	skips	x_hat + skips z + x_hat z only	Moderate:1
z only	2	2	4	0.50	skips x + x_hat + skips	x + skips z + x_hat	Minimal:1 Moderate:1	**3**	1	4	**0.75**	skips x + skips x + x_hat	x_hat + skips	Moderate:2 Weak:1
z + x	1	4	5	0.20	skips	x + skips x + x_hat z + skips z + x_hat	Moderate:1	1	1	2	0.50	skips	x_hat + skips	Strong:1
x + x_hat + skips	1	6	7	0.143	skips	x + skips x + x_hat x_hat + skips z + skips z + x_hat z only	Minimal:1	1	1	2	0.50	skips	x_hat + skips	Weak:1
skips	0	8	8	0.00	—	x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x z + x_hat z only	—	0	8	8	0.00	—	x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x z + x_hat z only	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 31: Practical evidence assessment of classifier tpye ablations in the P300 task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
eegnet_classifier__x	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:6	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:6
fc_classifier__z	**4**	1	5	**0.80**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x	Strong:3 Weak:1	0	5	5	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	—
fc_classifier__decoder_out	**4**	1	5	**0.80**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x	Strong:3 Moderate:1	**5**	1	6	**0.83**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat fc_classifier__z	eegnet_classifier__x	Strong:2 Weak:2 Moderate:1
eegnet_classifier__decoder_out	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Weak:2 Strong:1	2	2	4	0.50	eegnet_classifier__x_hat fc_classifier__z	eegnet_classifier__x fc_classifier__decoder_out	Strong:2
fc_classifier__x	1	4	5	0.20	eegnet_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:1	2	2	4	0.50	eegnet_classifier__x_hat fc_classifier__z	eegnet_classifier__x fc_classifier__decoder_out	Strong:2
fc_classifier__x_hat	1	4	5	0.20	eegnet_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:1	2	2	4	0.50	eegnet_classifier__x_hat fc_classifier__z	eegnet_classifier__x fc_classifier__decoder_out	Strong:2
eegnet_classifier__x_hat	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	0	5	5	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 32: Practical evidence assessment of architecture ablations in the P300 task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
no_ddpm__x__no_decoder	2	0	2	1.00	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:2	2	1	3	0.67	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:2
no_ddpm__x__use_decoder	2	0	2	1.00	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:2	2	1	3	0.67	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:2
use_ddpm__x__no_decoder	**3**	0	3	**1.00**	use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:2 Minimal:1	**5**	0	5	**1.00**	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Moderate:2 Strong:2 Weak:1
use_ddpm__x__use_decoder	2	1	3	0.67	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:2	2	1	3	0.67	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:2
use_ddpm__x_hat__no_decoder	1	4	5	0.20	use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Moderate:1	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__use_decoder	—
use_ddpm__x_hat__use_decoder	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder	—	1	4	5	0.20	use_ddpm__x_hat__no_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Moderate:1

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 33: Practical evidence assessment of mixup ablations in the P300 task (Part 1 of 2)

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
layer-1_randNo	6	0	6	1.00	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	—	Moderate:4, Minimal:1, Weak:1	**11**	0	11	**1.00**	layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Strong:9, Minimal:2
layer-1_randYes	5	0	5	1.00	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	—	Moderate:3, Weak:2	**10**	1	11	**0.91**	layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	layer-1_randNo	Strong:9, Minimal:1
layer0_randNo	**9**	0	9	**1.00**	layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Weak:8, Minimal:1	9	2	11	0.82	layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	layer-1_randNo, layer-1_randYes	Strong:6, Moderate:2, Weak:1
layer0_randYes	**6**	1	7	**0.86**	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randYes	layer0_randNo	Weak:5, Minimal:1	7	3	10	0.70	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo	Strong:4, Moderate:3
layer4_randNo	5	1	6	0.83	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	layer0_randNo	Weak:5	4	5	9	0.44	layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randYes	Moderate:3, Weak:1
layer1_randYes	5	2	7	0.71	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	layer-1_randNo, layer0_randNo	Moderate:5	0	11	11	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 34: Practical evidence assessment of mixup ablations in the P300 task (Part 2 of 2)

Config	Seen							Unseen						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
layer4_randYes	5	2	7	0.71	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	layer0_randNo, layer0_randYes	Moderate:3, Weak:2	7	3	10	0.70	layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo	Strong:5, Moderate:2
layer3_randYes	4	7	11	0.36	layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randYes, layer4_randNo, layer4_randYes	Weak:3, Minimal:1	1	5	6	0.17	layer1_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randYes	Weak:1
layer1_randNo	3	8	11	0.27	layer2_randNo, layer2_randYes, layer3_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randYes, layer3_randYes, layer4_randNo, layer4_randYes	Weak:2, Minimal:1	1	8	9	0.11	layer1_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer3_randNo, layer4_randNo, layer4_randYes	Moderate:1
layer2_randYes	1	9	10	0.10	layer2_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer3_randYes, layer4_randNo, layer4_randYes	Weak:1	4	5	9	0.44	layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer4_randYes	Moderate:1, Strong:1, Weak:1, Minimal:1
layer3_randNo	1	9	10	0.10	layer2_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer3_randYes, layer4_randNo, layer4_randYes	Minimal:1	2	7	9	0.22	layer1_randNo, layer1_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo, layer4_randYes	Weak:1, Moderate:1
layer2_randNo	0	11	11	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	1	7	8	0.13	layer1_randYes	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer2_randYes, layer4_randNo, layer4_randYes	Strong:1

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 35: Practical evidence assessment of decoder input ablations in the MI task

Config	Seen							Unseen						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
x_hat + skips	2	4	6	0.333	x + x_hat + skips z + skips	x + skips z + x z + x_hat z only	Weak:2	**5**	0	5	**1.00**	skips x + skips x + x_hat x + x_hat + skips z + skips z + x z + x_hat z only	—	Strong:5 Moderate:2 Weak:1
x + x_hat + skips	1	6	7	0.143	z + skips	x + skips x + x_hat x_hat + skips z + x z + x_hat z only	Minimal:1	2	5	7	0.286	skips z + skips	x + skips x + x_hat x_hat + skips z + x z + x_hat	Weak:1 Moderate:1
x + skips	3	2	5	0.60	x + x_hat + skips x_hat + skips z + skips	z + x z only	Weak:2 Moderate:1	**5**	0	5	**1.00**	skips x + x_hat + skips z + skips z + x_hat z only	—	Strong:2 Weak:1 Minimal:1 Moderate:1
z + skips	0	8	8	0.00	—	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + x z + x_hat z only	—	0	7	7	0.00	—	x + skips x + x_hat x + x_hat + skips x_hat + skips z + x z + x_hat z only	—
x + x_hat	3	3	6	0.50	skips x + x_hat + skips z + skips	z + x z + x_hat z only	Weak:2 Strong:1	4	0	4	1.00	skips x + x_hat + skips z + skips z only	—	Strong:2 Weak:1 Moderate:1
z only	**7**	1	8	**0.875**	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x_hat	z + x	Strong:4 Weak:3	2	4	6	0.333	skips z + skips	x + skips x + x_hat x_hat + skips z + x	Moderate:2
z + x_hat	5	2	7	0.714	skips x + x_hat x + x_hat + skips x_hat + skips z + skips	z + x z only	Strong:3 Weak:2	3	2	5	0.60	skips x + x_hat + skips z + skips	x + skips x_hat + skips	Strong:2 Weak:1
z + x	**8**	0	8	**1.00**	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x_hat z only	—	Strong:5 Moderate:3	4	0	4	1.00	skips x + x_hat + skips z + skips z only	—	Weak:2 Strong:2
skips	1	4	5	0.20	z + skips	x + x_hat z + x z + x_hat z only	Strong:1	0	7	7	0.00	—	x + skips x + x_hat x + x_hat + skips x_hat + skips z + x z + x_hat z only	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 36: Practical evidence assessment of classifier type ablations in the MI task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
eegnet_classifier__x	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:6	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:6
fc_classifier__z	**5**	1	6	**0.83**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x	Strong:5	**5**	1	6	**0.83**	eegnet_classifier__decoder_out eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x	Strong:5
eegnet_classifier__decoder_out	4	2	6	0.67	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x fc_classifier__z	Strong:3 Moderate:1	4	2	6	0.67	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	eegnet_classifier__x fc_classifier__z	Strong:3 Moderate:1
fc_classifier__decoder_out	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__z	Strong:3	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__z	Strong:3
fc_classifier__x	2	4	6	0.33	eegnet_classifier__x_hat fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:2	2	4	6	0.33	eegnet_classifier__x_hat fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__z	Strong:2
eegnet_classifier__x_hat	1	5	6	0.17	fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__z	Strong:1	1	5	6	0.17	fc_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__z	Strong:1
fc_classifier__x_hat	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__z	—	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__z	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 37: Practical evidence assessment of decoder input ablations in the MI task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
no_ddpm__x__use_decoder	**4**	0	4	**1.00**	no_ddpm__x__no_decoder use_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:3 Moderate:1	2	3	5	0.40	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Strong:2
use_ddpm__x__use_decoder	**4**	0	4	**1.00**	no_ddpm__x__no_decoder use_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:4	3	1	4	0.75	no_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:2 Weak:1
no_ddpm__x__no_decoder	2	2	4	0.50	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder use_ddpm__x__use_decoder	Strong:2	3	1	4	0.75	no_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	use_ddpm__x__no_decoder	Strong:3
use_ddpm__x__no_decoder	2	2	4	0.50	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder use_ddpm__x__use_decoder	Strong:2	**5**	0	5	**1.00**	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:3 Weak:1 Moderate:1
use_ddpm__x_hat__no_decoder	1	4	5	0.20	use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Strong:1	0	4	4	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	—
use_ddpm__x_hat__use_decoder	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder	—	0	4	4	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 38: Practical evidence assessment of mixup ablations in the MI task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
layer0_randYes	**10**	0	10	**1.00**	layer-1_randNo, layer-1_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Strong:7, Moderate:2, Minimal:1	**9**	0	9	**1.00**	layer-1_randNo, layer-1_randYes, layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo, layer4_randNo, layer4_randYes	—	Weak:6, Moderate:2, Strong:1
layer0_randNo	**9**	0	9	**1.00**	layer-1_randNo, layer-1_randYes, layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Moderate:6, Strong:3	1	3	4	0.25	layer1_randNo	layer0_randYes, layer2_randYes, layer3_randYes	Weak:1
layer1_randYes	**9**	1	10	**0.90**	layer-1_randNo, layer-1_randYes, layer1_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	layer0_randYes	Strong:6, Moderate:3	3	3	6	0.50	layer-1_randYes, layer1_randNo, layer4_randNo	layer0_randYes, layer2_randYes, layer3_randYes	Weak:3
layer3_randYes	6	3	9	0.67	layer-1_randNo, layer-1_randYes, layer2_randNo, layer2_randYes, layer4_randNo, layer4_randYes	layer0_randNo, layer0_randYes, layer1_randYes	Moderate:3, Strong:2, Weak:1	**8**	0	8	**1.00**	layer-1_randNo, layer-1_randYes, layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo, layer4_randNo	—	Weak:4, Moderate:4
layer3_randNo	5	3	8	0.63	layer-1_randNo, layer-1_randYes, layer2_randYes, layer4_randNo, layer4_randYes	layer0_randNo, layer0_randYes, layer1_randYes	Weak:3, Moderate:2	2	3	5	0.40	layer-1_randYes, layer1_randNo	layer0_randYes, layer2_randYes, layer3_randYes	Minimal:1, Weak:1
layer1_randNo	4	3	7	0.57	layer-1_randNo, layer2_randYes, layer4_randNo, layer4_randYes	layer0_randNo, layer0_randYes, layer1_randYes	Moderate:2, Weak:1, Minimal:1	0	9	9	0.00	—	layer-1_randNo, layer0_randNo, layer0_randYes, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randYes	—
layer2_randNo	3	4	7	0.43	layer-1_randNo, layer2_randYes, layer4_randNo	layer0_randNo, layer0_randYes, layer1_randYes, layer3_randYes	Weak:1, Minimal:1, Moderate:1	3	3	6	0.50	layer-1_randNo, layer-1_randYes, layer1_randNo	layer0_randYes, layer2_randYes, layer3_randYes	Moderate:2, Minimal:1
layer-1_randYes	1	5	6	0.17	layer-1_randNo	layer0_randNo, layer0_randYes, layer1_randYes, layer3_randNo, layer3_randYes	Weak:1	0	8	8	0.00	—	layer-1_randNo, layer0_randYes, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randYes	—
layer2_randYes	1	7	8	0.13	layer4_randNo	layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo, layer3_randYes	Weak:1	**8**	0	8	**1.00**	layer-1_randNo, layer-1_randYes, layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo, layer4_randYes	—	Weak:4, Strong:2, Moderate:1, Minimal:1
layer-1_randNo	0	8	8	0.00	—	layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo, layer3_randYes	—	**2**	4	6	**0.33**	layer-1_randYes, layer1_randNo	layer0_randYes, layer2_randNo, layer2_randYes, layer3_randYes	Weak:1, Moderate:1
layer4_randNo	0	8	8	0.00	—	layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes	—	0	3	3	0.00	—	layer0_randYes, layer1_randYes, layer3_randYes	—
layer4_randYes	0	6	6	0.00	—	layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer3_randNo, layer3_randYes	—	**2**	2	4	**0.50**	layer-1_randYes, layer1_randNo	layer0_randYes, layer2_randYes	Weak:2

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 39: Practical evidence assessment of decoder input ablations in the imagined speech task

Config	Seen							Unseen						
	Wins	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**	**Wins**	**Losses**	**Total**	**Rate**	**Win List**	**Loss List**	**Evidence**
x + x_hat	**5**	0	5	**1.00**	skips x + skips z + skips z + x z only	—	Moderate:3 Strong:1 Weak:1	1	7	8	0.125	z + skips	skips x + skips x + x_hat + skips x_hat + skips z + x z + x_hat z only	Weak:1
z + x	**5**	1	6	**0.833**	x + skips x + x_hat + skips x_hat + skips z + x_hat z only	x + x_hat	Weak:4 Strong:1	4	0	4	1.00	skips x + x_hat + skips z + skips z only	—	Weak:2 Strong:2
z + skips	4	1	5	0.800	x + skips x + x_hat + skips x_hat + skips z only	x + x_hat	Weak:2 Moderate:2	0	7	7	0.000	—	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + x z + x_hat z only	—
skips	3	1	4	0.750	x + skips x + x_hat + skips z only	x + x_hat	Weak:2 Moderate:1	3	5	8	0.375	x + x_hat z + skips z + x	x + skips x + x_hat + skips x_hat + skips z + x_hat z only	Weak:3
z + x_hat	3	1	4	0.750	x + x_hat + skips x_hat + skips z only	z + x	Weak:2 Strong:1	**8**	0	8	**1.00**	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x z only	—	Weak:4 Moderate:3 Strong:1
x_hat + skips	1	3	4	0.250	z only	z + skips z + x z + x_hat	Strong:1	**6**	1	7	**0.857**	skips x + skips x + x_hat z + skips z + x z only	z + x_hat	Weak:3 Moderate:2 Strong:1
x + skips	1	4	5	0.200	z only	skips x + x_hat z + skips z + x	Moderate:1	3	3	6	0.500	skips x + x_hat z + skips	x + x_hat + skips x_hat + skips z + x_hat	Weak:1 Strong:1 Moderate:1
x + x_hat + skips	1	4	5	0.200	z only	skips z + skips z + x z + x_hat	Moderate:1	**6**	1	7	**0.857**	skips x + skips x + x_hat z + skips z + x z only	z + x_hat	Weak:3 Moderate:2 Strong:1
z only	0	8	8	0.00	—	skips x + skips x + x_hat x + x_hat + skips x_hat + skips z + skips z + x z + x_hat	—	4	3	7	0.571	skips x + x_hat z + skips z + x	x + x_hat + skips x_hat + skips z + x_hat	Weak:3 Moderate:1

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 40: Practical evidence assessment of classifier type ablations in the imagined speech task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
fc_classifier__decoder_out	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	Strong:3 Moderate:3	1	4	5	0.20	eegnet_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__x_hat fc_classifier__z	Weak:1
fc_classifier__x	**4**	1	5	**0.80**	eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__x_hat fc_classifier__z	fc_classifier__decoder_out	Strong:3 Weak:1	0	4	4	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__x_hat fc_classifier__z	—
eegnet_classifier__decoder_out	**4**	1	5	**0.80**	eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__x_hat fc_classifier__z	fc_classifier__decoder_out	Strong:3 Moderate:1	**4**	1	5	**0.80**	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	fc_classifier__z	Moderate:3 Strong:1
fc_classifier__z	3	3	6	0.50	eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__x_hat	eegnet_classifier__decoder_out fc_classifier__decoder_out fc_classifier__x	Strong:3	**6**	0	6	**1.00**	eegnet_classifier__decoder_out eegnet_classifier__x eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	—	Strong:3 Moderate:2 Weak:1
eegnet_classifier__x	2	4	6	0.33	eegnet_classifier__x_hat fc_classifier__x_hat	eegnet_classifier__decoder_out fc_classifier__decoder_out fc_classifier__x fc_classifier__z	Weak:2	**4**	1	5	**0.80**	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat	fc_classifier__z	Strong:2 Moderate:2
fc_classifier__x_hat	1	5	6	0.17	eegnet_classifier__x_hat	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__z	Weak:1	3	3	6	0.50	eegnet_classifier__x_hat fc_classifier__decoder_out fc_classifier__x	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__z	Weak:2 Strong:1
eegnet_classifier__x_hat	0	6	6	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x fc_classifier__x_hat fc_classifier__z	—	0	5	5	0.00	—	eegnet_classifier__decoder_out eegnet_classifier__x fc_classifier__decoder_out fc_classifier__x_hat fc_classifier__z	—

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 41: Practical evidence assessment of architecture ablations in the imagined speech task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
no_ddpm__x__no_decoder	**4**	0	4	**1.00**	use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:3 Weak:1	3	2	5	0.60	use_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder use_ddpm__x__use_decoder	Moderate:3
no_ddpm__x__use_decoder	3	0	3	1.00	use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:2 Moderate:1	**5**	0	5	**1.00**	no_ddpm__x__no_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	—	Strong:4 Moderate:1
use_ddpm__x__no_decoder	3	1	4	0.75	use_ddpm__x__use_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder	Strong:3	2	2	4	0.50	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder	Weak:1 Moderate:1
use_ddpm__x__use_decoder	2	3	5	0.40	use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder	Strong:2	3	1	4	0.75	no_ddpm__x__no_decoder use_ddpm__x_hat__no_decoder use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder	Moderate:2 Weak:1
use_ddpm__x_hat__no_decoder	0	4	4	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	—	0	5	5	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder use_ddpm__x_hat__use_decoder	—
use_ddpm__x_hat__use_decoder	0	4	4	0.00	—	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	—	1	4	5	0.20	use_ddpm__x_hat__no_decoder	no_ddpm__x__no_decoder no_ddpm__x__use_decoder use_ddpm__x__no_decoder use_ddpm__x__use_decoder	Weak:1

Each configuration's wins, losses, total comparisons, and win rate are reported, along with the corresponding lists of winning and losing opponents and an evidence summary.

Table 42: Practical evidence assessment of mixup ablations in the imagined speech task

Config	Seen							Unseen						
	Wins	Losses	Total	Rate	Win List	Loss List	Evidence	Wins	Losses	Total	Rate	Win List	Loss List	Evidence
layer0_randYes	**11**	0	11	**1.00**	layer-1_randNo, layer-1_randYes, layer0_randNo, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo, layer4_randYes	—	Strong:11	**5**	2	7	**0.71**	layer-1_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	layer-1_randNo, layer2_randYes	Strong:3, Weak:2
layer1_randNo	**7**	1	8	**0.88**	layer-1_randNo, layer-1_randYes, layer0_randNo, layer2_randYes, layer3_randNo, layer4_randNo, layer4_randYes	layer0_randYes	Weak:6, Moderate:1	4	6	10	0.40	layer-1_randYes, layer1_randYes, layer2_randNo, layer3_randNo	layer-1_randNo, layer0_randYes, layer2_randYes, layer3_randYes, layer4_randNo, layer4_randYes	Weak:2, Strong:1, Moderate:1
layer1_randYes	6	1	7	0.86	layer0_randNo, layer2_randNo, layer2_randYes, layer3_randNo, layer3_randYes, layer4_randNo	layer0_randYes	Moderate:6	1	3	4	0.25	layer3_randNo	layer0_randYes, layer1_randNo, layer2_randYes	Moderate:1
layer2_randNo	3	2	5	0.60	layer0_randNo, layer2_randYes, layer4_randNo	layer0_randYes, layer1_randYes	Moderate:2, Weak:1	1	9	10	0.10	layer3_randNo	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer2_randYes, layer3_randYes, layer4_randNo, layer4_randYes	Strong:1
layer3_randNo	4	3	7	0.57	layer0_randNo, layer2_randYes, layer3_randYes, layer4_randNo	layer0_randYes, layer1_randNo, layer1_randYes	Moderate:2, Weak:1, Strong:1	0	11	11	0.00	—	layer-1_randNo, layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer2_randYes, layer3_randYes, layer4_randNo, layer4_randYes	—
layer-1_randNo	1	2	3	0.33	layer4_randNo	layer0_randYes, layer1_randNo	Moderate:1	**6**	0	6	**1.00**	layer-1_randYes, layer0_randNo, layer0_randYes, layer1_randNo, layer2_randNo, layer3_randNo	—	Moderate:3, Weak:2, Strong:1
layer0_randNo	1	5	6	0.17	layer4_randNo	layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	Moderate:1	3	1	4	0.75	layer-1_randYes, layer2_randNo, layer3_randNo	layer-1_randNo	Moderate:2, Strong:1
layer-1_randYes	0	2	2	0.00	—	layer0_randYes, layer1_randNo	—	2	4	6	0.33	layer2_randNo, layer3_randNo	layer-1_randNo, layer0_randNo, layer0_randYes, layer1_randNo	Moderate:1, Strong:1
layer2_randYes	0	5	5	0.00	—	layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	—	5	0	5	1.00	layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	—	Moderate:3, Weak:1, Strong:1
layer3_randYes	0	3	3	0.00	—	layer0_randYes, layer1_randYes, layer3_randNo	—	3	0	3	1.00	layer1_randNo, layer2_randNo, layer3_randNo	—	Weak:1, Moderate:1, Strong:1
layer4_randNo	0	7	7	0.00	—	layer-1_randNo, layer0_randNo, layer0_randYes, layer1_randNo, layer1_randYes, layer2_randNo, layer3_randNo	—	3	0	3	1.00	layer1_randNo, layer2_randNo, layer3_randNo	—	Weak:1, Moderate:1, Strong:1
layer4_randYes	0	2	2	0.00	—	layer0_randYes, layer1_randNo	—	3	0	3	1.00	layer1_randNo, layer2_randNo, layer3_randNo	—	Weak:1, Moderate:1, Strong:1

Each row shows win/loss stats and qualitative evidence for seen and unseen subject comparisons.

Table 43: Bayesian evidence assessment of decoder input ablations on seen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
x + x_hat + skips	x + x_hat	0.006993006993	0.4072213877	0.4576239803	0.135154632	3
x + x_hat + skips	x_hat + skips	0.001554001554	0.1277370191	0.7904370663	0.08182591465	3
x + x_hat + skips	x + skips	0.002331002331	0.05622746549	0.9195907337	0.02418180077	3
x + x_hat + skips	skips	0.01825951826	0.6552695743	0.2173670696	0.1273633562	3
x + x_hat + skips	z only	0.00777000777	0.4601924043	0.3082698215	0.2315377742	3
x + x_hat + skips	z + x	0.009712509713	0.4872757353	0.4465708069	0.06615345775	3
x + x_hat + skips	z + x_hat	0.006993006993	0.291618297	0.6748438492	0.03353785374	3
x + x_hat + skips	z + skips	0.005439005439	0.3449252349	0.5256993868	0.1293753783	3
x + x_hat	x_hat + skips	-0.005439005439	0.165056371	0.463719159	0.37122447	3
x + x_hat	x + skips	-0.004662004662	0.1867724085	0.4536427293	0.3595848622	3
x + x_hat	skips	0.01126651127	0.5451105314	0.3726927806	0.08219668798	3
x + x_hat	z only	0.000777000777	0.2351467355	0.5595924463	0.2052608182	3
x + x_hat	z + x	0.00271950272	0.1342152618	0.8069448934	0.05883984476	3
x + x_hat	z + x_hat	3.70E-17	0.2621371795	0.475725641	0.2621371795	3
x + x_hat	z + skips	-0.001554001554	0.2788090622	0.3907941262	0.3303968116	3
x_hat + skips	x + skips	0.000777000777	0.057196689	0.8989675316	0.0438357794	3
x_hat + skips	skips	0.01670551671	0.6462311243	0.2402124945	0.1135563811	3
x_hat + skips	z only	0.006216006216	0.4259897872	0.3439535126	0.2300567002	3
x_hat + skips	z + x	0.008158508159	0.4138874767	0.5186213571	0.06749116622	3
x_hat + skips	z + x_hat	0.005439005439	0.05348962787	0.9410602535	0.005450118599	3
x_hat + skips	z + skips	0.003885003885	0.1739873248	0.7710487483	0.05496392697	3
x + skips	skips	0.01592851593	0.6128314643	0.2430106582	0.1441578775	3
x + skips	z only	0.005439005439	0.4213678234	0.3158824392	0.2627497374	3
x + skips	z + x	0.007381507382	0.3977108733	0.5079004852	0.09438864146	3
x + skips	z + x_hat	0.004662004662	0.0644456337	0.9253444009	0.01020996542	3
x + skips	z + skips	0.003108003108	0.2208576006	0.673134884	0.1060075154	3
skips	z only	-0.01048951049	0.009178275343	0.4297996164	0.5610221082	3
skips	z + x	-0.008547008547	0.1032182657	0.4476016877	0.4491800465	3
skips	z + x_hat	-0.01126651127	0.1694595733	0.304340691	0.5261997357	3
skips	z + skips	-0.01282051282	0.1632223436	0.2808168648	0.5559607915	3
z only	z + x	0.001942501943	0.2821134095	0.5093980036	0.208488587	3
z only	z + x_hat	-0.000777000777	0.3171595098	0.3421932156	0.3406472746	3
z only	z + skips	-0.002331002331	0.3033447828	0.3252097791	0.3714454381	3
z + x	z + x_hat	-0.00271950272	0.1386152854	0.618199462	0.2431852526	3
z + x	z + skips	-0.004273504274	0.1744886081	0.4881975089	0.337313883	3
z + x_hat	z + skips	-0.001554001554	0.0793232897	0.7962580914	0.1244186189	3

Table 44: Bayesian evidence assessment of decoder input ablations on unseen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
x + x_hat + skips	x + x_hat	-0.01282051282	0.0491903225	0.3260644706	0.6247452069	3
x + x_hat + skips	x_hat + skips	-0.004273504274	0.1753950877	0.4865895812	0.3380153311	3
x + x_hat + skips	x + skips	0.01388888889	0.7620140193	0.2213532638	0.01663271685	3
x + x_hat + skips	skips	0.005341880342	0.4343899953	0.265424077	0.3001859277	3
x + x_hat + skips	z only	-0.01602564103	0.1639164971	0.233235663	0.6028478399	3
x + x_hat + skips	z + x	0.01068376068	0.5356858485	0.4182145309	0.04609962059	3
x + x_hat + skips	z + x_hat	0.005341880342	0.4334799737	0.2686965897	0.2978234365	3
x + x_hat + skips	z + skips	-0.004273504274	0.2210581421	0.4091356847	0.3698061732	3
x + x_hat	x_hat + skips	0.008547008547	0.4736673258	0.3055536853	0.2207789889	3
x + x_hat	x + skips	0.02670940171	0.8481200192	0.1045169345	0.04736304634	3
x + x_hat	skips	0.01816239316	0.6595313467	0.2193905632	0.1210780901	3
x + x_hat	z only	-0.003205128205	0.2513575889	0.3900876037	0.3585548074	3
x + x_hat	z + x	0.0235042735	0.9215492218	0.06270373278	0.01574704541	3
x + x_hat	z + x_hat	0.01816239316	0.5995193453	0.1874197081	0.2130609465	3
x + x_hat	z + skips	0.008547008547	0.4309028719	0.5050801286	0.06401699946	3
x_hat + skips	x + skips	0.01816239316	0.8084563537	0.1605116455	0.03103200074	3
x_hat + skips	skips	0.009615384615	0.496279061	0.1811444739	0.3225764651	3
x_hat + skips	z only	-0.01175213675	0.2745850211	0.2050900957	0.5203248832	3
x_hat + skips	z + x	0.01495726496	0.5974712829	0.256193859	0.1463348581	3
x_hat + skips	z + x_hat	0.009615384615	0.4945607409	0.248020708	0.2574185511	3
x_hat + skips	z + skips	0	0.3725136154	0.2549727692	0.3725136154	3
x + skips	skips	-0.008547008547	0.2957714672	0.2217461067	0.4824824261	3
x + skips	z only	-0.02991452991	0.1179607274	0.1276686708	0.7543706018	3
x + skips	z + x	-0.003205128205	0.1686075062	0.5386346688	0.292757825	3
x + skips	z + x_hat	-0.008547008547	0.2589232622	0.2626141784	0.4784625594	3
x + skips	z + skips	-0.01816239316	0.1425873074	0.2153816627	0.6420310299	3
skips	z only	-0.02136752137	0.1243311617	0.1849632848	0.6907055535	3
skips	z + x	0.005341880342	0.4231177869	0.3049383138	0.2719438993	3
skips	z + x_hat	-3.70E-17	0.4155708997	0.1688582007	0.4155708997	3
skips	z + skips	-0.009615384615	0.09629904233	0.4171125382	0.4865884194	3
z only	z + x	0.02670940171	0.8291035932	0.1143919903	0.05650441656	3
z only	z + x_hat	0.02136752137	0.6598087955	0.1808495094	0.159341695	3
z only	z + skips	0.01175213675	0.5446226142	0.3272150548	0.1281623311	3
z + x	z + x_hat	-0.005341880342	0.281641804	0.2912679113	0.4270902847	3
z + x	z + skips	-0.01495726496	0.0765139553	0.2726360719	0.6508499728	3
z + x_hat	z + skips	-0.009615384615	0.3021151679	0.202109754	0.4957750781	3

Table 45: Bayesian evidence assessment of classifier type ablations on seen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
eegnet_classifier__x	eegnet_classifier__x_hat	0.7855477855	0.9974747527	0.0001244714922	0.002400775803	3
eegnet_classifier__x	eegnet_classifier__decoder_out	0.08935508936	0.9978271584	0.0007834572033	0.001389384359	3
eegnet_classifier__x	fc_classifier__x	0.7843822844	0.9996831675	1.57E-05	0.0003010938821	3
eegnet_classifier__x	fc_classifier__x_hat	0.8543123543	0.999946812	2.43E-06	5.08E-05	3
eegnet_classifier__x	fc_classifier__decoder_out	0.1153846154	0.9850522476	0.004248642321	0.0106991101	3
eegnet_classifier__x	fc_classifier__z	0.04079254079	0.9479425423	0.0308853512	0.02117210646	3
eegnet_classifier__x_hat	eegnet_classifier__decoder_out	-0.6961926962	0.003419689093	0.0002000567375	0.9963802542	3
eegnet_classifier__x_hat	fc_classifier__x	-0.001165501166	0.3978407016	0.1836293954	0.4185299031	3
eegnet_classifier__x_hat	fc_classifier__x_hat	0.06876456876	0.8055549637	0.05399336546	0.1404516708	3
eegnet_classifier__x_hat	fc_classifier__decoder_out	-0.6701631702	0.005419803428	0.0003276395382	0.994252557	3
eegnet_classifier__x_hat	fc_classifier__z	-0.7447552448	0.002980693126	0.0001629291043	0.9968563778	3
eegnet_classifier__decoder_out	fc_classifier__x	0.695027195	0.9995034081	2.77E-05	0.0004688562281	3
eegnet_classifier__decoder_out	fc_classifier__x_hat	0.764957265	0.9998358972	8.36E-06	0.0001557457272	3
eegnet_classifier__decoder_out	fc_classifier__decoder_out	0.02602952603	0.7551509035	0.1448922333	0.09995686318	3
eegnet_classifier__decoder_out	fc_classifier__z	-0.04856254856	0.00604968638	0.007581437855	0.9863688758	3
fc_classifier__x	fc_classifier__x_hat	0.06993006993	0.9425081197	0.02261977183	0.03487210851	3
fc_classifier__x	fc_classifier__decoder_out	-0.668997669	0.001548222681	9.49E-05	0.9983568467	3
fc_classifier__x	fc_classifier__z	-0.7435897436	0.0003773041956	2.08E-05	0.9996018672	3
fc_classifier__x_hat	fc_classifier__decoder_out	-0.7389277389	0.0002752569251	1.53E-05	0.9997094444	3
fc_classifier__x_hat	fc_classifier__z	-0.8135198135	0.0002801819742	1.41E-05	0.9997057092	3
fc_classifier__decoder_out	fc_classifier__z	-0.07459207459	0.03847802095	0.02264037834	0.9388816007	3

Table 46: Bayesian evidence assessment of classifier variants on unseen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
eegnet_classifier__x	eegnet_classifier__x_hat	0.719017094	0.9941250768	0.0003126409919	0.005562282177	3
eegnet_classifier__x	eegnet_classifier__decoder_out	0.1047008547	0.8517898021	0.03221550818	0.1159946897	3
eegnet_classifier__x	fc_classifier__x	0.6965811966	0.9941798993	0.0003194766474	0.005500624028	3
eegnet_classifier__x	fc_classifier__x_hat	0.7702991453	0.9919474317	0.0003981951421	0.007654373153	3
eegnet_classifier__x	fc_classifier__decoder_out	0.1153846154	0.9031687756	0.02244695533	0.0743842691	3
eegnet_classifier__x	fc_classifier__z	-0.03632478632	0.1578538502	0.1068760593	0.7352700904	3
eegnet_classifier__x_hat	eegnet_classifier__decoder_out	-0.6143162393	0.0003327796059	2.24E-05	0.9996448529	3
eegnet_classifier__x_hat	fc_classifier__x	-0.02243589744	0.1644621354	0.171609785	0.6639280796	3
eegnet_classifier__x_hat	fc_classifier__x_hat	0.05128205128	0.6739337523	0.06729475814	0.2587714896	3
eegnet_classifier__x_hat	fc_classifier__decoder_out	-0.6036324786	0.000797597804	5.45E-05	0.9991478929	3
eegnet_classifier__x_hat	fc_classifier__z	-0.7553418803	0.00335896058	0.0001807472277	0.9964602922	3
eegnet_classifier__decoder_out	fc_classifier__x	0.5918803419	0.9997431462	1.68E-05	0.000240079429	3
eegnet_classifier__decoder_out	fc_classifier__x_hat	0.6655982906	0.9951195884	0.0002807606336	0.004599650926	3
eegnet_classifier__decoder_out	fc_classifier__decoder_out	0.01068376068	0.5183108555	0.3529521795	0.1287369649	3
eegnet_classifier__decoder_out	fc_classifier__z	-0.141025641	0.04449883101	0.01221778853	0.9432833805	3
fc_classifier__x	fc_classifier__x_hat	0.07371794872	0.8224598002	0.0488305597	0.1287096401	3
fc_classifier__x	fc_classifier__decoder_out	-0.5811965812	0.0007981961578	5.67E-05	0.9991450746	3
fc_classifier__x	fc_classifier__z	-0.7329059829	0.002228941662	0.0001241570405	0.9976469013	3
fc_classifier__x_hat	fc_classifier__decoder_out	-0.6549145299	0.005394449501	0.0003339544716	0.994271596	3
fc_classifier__x_hat	fc_classifier__z	-0.8066239316	0.002928727096	0.0001475315086	0.9969237414	3
fc_classifier__decoder_out	fc_classifier__z	-0.1517094017	0.02672121585	0.007268100357	0.9660106838	3

Table 47: Bayesian evidence assessment of architecture ablations on seen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
use_ddpm__x__use_decoder	use_ddpm__x__no_decoder	-0.004273504274	0.1467848185	0.539083023	0.3141321585	3
use_ddpm__x__use_decoder	use_ddpm__x_hat__use_decoder	0.7855477855	0.9999548373	2.24E-06	4.29E-05	3
use_ddpm__x__use_decoder	use_ddpm__x_hat__no_decoder	0.7836052836	0.999244801	3.75E-05	0.0007176953454	3
use_ddpm__x__use_decoder	no_ddpm__x__use_decoder	-0.05205905206	0.03269113195	0.03133239227	0.9359764758	3
use_ddpm__x__use_decoder	no_ddpm__x__no_decoder	-0.05866355866	0.05146770661	0.03796571886	0.9105665745	3
use_ddpm__x__no_decoder	use_ddpm__x_hat__use_decoder	0.7898212898	0.9999754926	1.21E-06	2.33E-05	3
use_ddpm__x__no_decoder	use_ddpm__x_hat__no_decoder	0.7878787879	0.9996523997	1.72E-05	0.0003304095815	3
use_ddpm__x__no_decoder	no_ddpm__x__use_decoder	-0.04778554779	0.008104232996	0.01025732533	0.9816384417	3
use_ddpm__x__no_decoder	no_ddpm__x__no_decoder	-0.05439005439	0.02293724067	0.02198948784	0.9550732715	3
use_ddpm__x_hat__use_decoder	use_ddpm__x_hat__no_decoder	-0.001942501943	0.3284789163	0.2911427916	0.380378292	3
use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder	-0.8376068376	0.0001152440137	5.64E-06	0.9998791207	3
use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder	-0.8442113442	0.0001994892911	9.67E-06	0.9997908367	3
use_ddpm__x_hat__no_decoder	no_ddpm__x__use_decoder	-0.8356643357	0.0002238272147	1.10E-05	0.9997652057	3
use_ddpm__x_hat__no_decoder	no_ddpm__x__no_decoder	-0.8422688423	2.68E-05	1.30E-06	0.9999719478	3
no_ddpm__x__use_decoder	no_ddpm__x__no_decoder	-0.006604506605	0.1552197591	0.4403376075	0.4044426334	3

Table 48: Bayesian evidence assessment of architecture ablations for unSeen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
use_ddpm__x__use_decoder	use_ddpm__x__no_decoder	0.01175213675	0.5725664853	0.3657018961	0.06173161858	3
use_ddpm__x__use_decoder	use_ddpm__x_hat__use_decoder	0.7713675214	0.9975348705	0.0001237069014	0.002341422619	3
use_ddpm__x__use_decoder	use_ddpm__x_hat__no_decoder	0.7767094017	0.996752828	0.0001614971039	0.003085674931	3
use_ddpm__x__use_decoder	no_ddpm__x__use_decoder	-0.002136752137	0.3462081736	0.2562844925	0.3975073338	3
use_ddpm__x__use_decoder	no_ddpm__x__no_decoder	-0.008547008547	0.01580831276	0.6296662858	0.3545254015	3
use_ddpm__x__no_decoder	use_ddpm__x_hat__use_decoder	0.7596153846	0.9976326959	0.0001206223479	0.002246681788	3
use_ddpm__x__no_decoder	use_ddpm__x_hat__no_decoder	0.764957265	0.9969466315	0.0001542178536	0.002899150601	3
use_ddpm__x__no_decoder	no_ddpm__x__use_decoder	-0.01388888889	0.2474675827	0.2051155102	0.5474169071	3
use_ddpm__x__no_decoder	no_ddpm__x__no_decoder	-0.0202991453	0.02714667705	0.1212528557	0.8516004673	3
use_ddpm__x_hat__use_decoder	use_ddpm__x_hat__no_decoder	0.005341880342	0.347373062	0.5156378604	0.1369890776	3
use_ddpm__x_hat__use_decoder	no_ddpm__x__use_decoder	-0.7735042735	0.005124050777	0.0002675815198	0.9946083677	3
use_ddpm__x_hat__use_decoder	no_ddpm__x__no_decoder	-0.7799145299	0.002494230703	0.0001302361551	0.9973755331	3
use_ddpm__x_hat__no_decoder	no_ddpm__x__use_decoder	-0.7788461538	0.006097907255	0.0003152188219	0.9935868739	3
use_ddpm__x_hat__no_decoder	no_ddpm__x__no_decoder	-0.7852564103	0.003223991823	0.0001667800573	0.9966092281	3
no_ddpm__x__use_decoder	no_ddpm__x__no_decoder	-0.00641025641	0.281984361	0.2707144165	0.4473012224	3

Table 49: Bayesian evidence assessment of classifier type ablations on unseen subjects in the SSVEP task

config_1	config_2	mean_diff	p_left	p_rope	p_right	n_seeds
eegnet_classifier__x	eegnet_classifier__x_hat	0.719017094	0.9941250768	0.0003126409919	0.005562282177	3
eegnet_classifier__x	eegnet_classifier__decoder_out	0.1047008547	0.8517898021	0.03221550818	0.1159946897	3
eegnet_classifier__x	fc_classifier__x	0.6965811966	0.9941798993	0.0003194766474	0.005500624028	3
eegnet_classifier__x	fc_classifier__x_hat	0.7702991453	0.9919474317	0.0003981951421	0.007654373153	3
eegnet_classifier__x	fc_classifier__decoder_out	0.1153846154	0.9031687756	0.02244695533	0.0743842691	3
eegnet_classifier__x	fc_classifier__z	-0.03632478632	0.1578538502	0.1068760593	0.7352700904	3
eegnet_classifier__x_hat	eegnet_classifier__decoder_out	-0.6143162393	0.0003327796059	2.24E-05	0.9996448529	3
eegnet_classifier__x_hat	fc_classifier__x	-0.02243589744	0.1644621354	0.171609785	0.6639280796	3
eegnet_classifier__x_hat	fc_classifier__x_hat	0.05128205128	0.6739337523	0.06729475814	0.2587714896	3
eegnet_classifier__x_hat	fc_classifier__decoder_out	-0.6036324786	0.000797597804	5.45E-05	0.9991478929	3
eegnet_classifier__x_hat	fc_classifier__z	-0.7553418803	0.00335896058	0.0001807472277	0.9964602922	3
eegnet_classifier__decoder_out	fc_classifier__x	0.5918803419	0.9997431462	1.68E-05	0.000240079429	3
eegnet_classifier__decoder_out	fc_classifier__x_hat	0.6655982906	0.9951195884	0.0002807606336	0.004599650926	3
eegnet_classifier__decoder_out	fc_classifier__decoder_out	0.01068376068	0.5183108555	0.3529521795	0.1287369649	3
eegnet_classifier__decoder_out	fc_classifier__z	-0.141025641	0.04449883101	0.01221778853	0.9432833805	3
fc_classifier__x	fc_classifier__x_hat	0.07371794872	0.8224598002	0.0488305597	0.1287096401	3
fc_classifier__x	fc_classifier__decoder_out	-0.5811965812	0.0007981961578	5.67E-05	0.9991450746	3
fc_classifier__x	fc_classifier__z	-0.7329059829	0.002228941662	0.0001241570405	0.9976469013	3
fc_classifier__x_hat	fc_classifier__decoder_out	-0.6549145299	0.005394449501	0.0003339544716	0.994271596	3
fc_classifier__x_hat	fc_classifier__z	-0.8066239316	0.002928727096	0.0001475315086	0.9969237414	3
fc_classifier__decoder_out	fc_classifier__z	-0.1517094017	0.02672121585	0.007268100357	0.9660106838	3

Table 50: Pairwise Wilcoxon test results for decoder input ablations on seen subjects in the SSVEP task

config_1	config_2	mean_diff	p_value	mean_1	mean_2	n_seeds
x + x_hat + skips	x + x_hat	0.0069930070	0.5	0.8586	0.8516	3
x + x_hat + skips	x_hat + skips	0.0015540016	0.75	0.8586	0.8570	3
x + x_hat + skips	x + skips	0.0023310023	0.5	0.8586	0.8563	3
x + x_hat + skips	skips	0.0182595183	0.25	0.8586	0.8403	3
x + x_hat + skips	z only	0.0077700078	0.75	0.8586	0.8508	3
x + x_hat + skips	z + x	0.0097125097	0.25	0.8586	0.8489	3
x + x_hat + skips	z + x_hat	0.0069930070	0.25	0.8586	0.8516	3
x + x_hat + skips	z + skips	0.0054380054	0.5	0.8586	0.8531	3
x + x_hat	x_hat + skips	-0.0054380054	0.75	0.8516	0.8570	3
x + x_hat	x + skips	-0.0046610047	0.5	0.8516	0.8563	3
x + x_hat	skips	0.0112665113	0.25	0.8516	0.8403	3
x + x_hat	z only	0.0007770008	1.0	0.8516	0.8508	3
x + x_hat	z + x	0.0027195027	0.5	0.8516	0.8489	3
x + x_hat	z + x_hat	0.0000000000	1.0	0.8516	0.8516	3
x + x_hat	z + skips	-0.0015540016	0.75	0.8516	0.8531	3
x_hat + skips	x + skips	0.0007770008	1.0	0.8570	0.8563	3
x_hat + skips	skips	0.0167055167	0.5	0.8570	0.8403	3
x_hat + skips	z only	0.0062160062	0.75	0.8570	0.8508	3
x_hat + skips	z + x	0.0081585082	0.5	0.8570	0.8489	3
x_hat + skips	z + x_hat	0.0054380054	0.25	0.8570	0.8516	3
x_hat + skips	z + skips	0.0038850039	0.5	0.8570	0.8531	3
x + skips	skips	0.0159285159	0.5	0.8563	0.8403	3
x + skips	z only	0.0054380054	0.75	0.8563	0.8508	3
x + skips	z + x	0.0073815074	0.5	0.8563	0.8489	3
x + skips	z + x_hat	0.0046610047	0.25	0.8563	0.8516	3
x + skips	z + skips	0.0031070031	0.75	0.8563	0.8531	3
skips	z only	-0.0104895105	0.25	0.8403	0.8508	3
skips	z + x	-0.0085470085	0.5	0.8403	0.8489	3
skips	z + x_hat	-0.0112665113	0.5	0.8403	0.8516	3
skips	z + skips	-0.0128205128	0.5	0.8403	0.8531	3
z only	z + x	0.0019425019	0.75	0.8508	0.8489	3
z only	z + x_hat	-0.0007770008	1.0	0.8508	0.8516	3
z only	z + skips	-0.0023310023	1.0	0.8508	0.8531	3
z + x	z + x_hat	-0.0027195027	1.0	0.8489	0.8516	3
z + x	z + skips	-0.0042735043	0.75	0.8489	0.8531	3
z + x_hat	z + skips	-0.0015540016	1.0	0.8516	0.8531	3

Table 51: Pairwise Wilcoxon test results for decoder input ablations on unseen subjects in the SSVEP task

config_1	config_2	mean_diff	p_value	mean_1	mean_2	n_seeds
x + x_hat + skips	x + x_hat	-0.0128205128	0.25	0.8312	0.8440	3
x + x_hat + skips	x_hat + skips	-0.0042735043	0.75	0.8312	0.8355	3
x + x_hat + skips	x + skips	0.0138888889	0.25	0.8312	0.8173	3
x + x_hat + skips	skips	0.0053418803	1.0	0.8312	0.8259	3
x + x_hat + skips	z only	-0.0160256410	0.5	0.8312	0.8472	3
x + x_hat + skips	z + x	0.0106837607	0.25	0.8312	0.8205	3
x + x_hat + skips	z + x_hat	0.0053418803	1.0	0.8312	0.8259	3
x + x_hat + skips	z + skips	-0.0042735043	1.0	0.8312	0.8355	3
x + x_hat	x_hat + skips	0.0085470085	0.75	0.8440	0.8355	3
x + x_hat	x + skips	0.0267094017	0.25	0.8440	0.8173	3
x + x_hat	skips	0.0181623932	0.5	0.8440	0.8259	3
x + x_hat	z only	-0.0032051282	0.75	0.8440	0.8472	3
x + x_hat	z + x	0.0235042735	0.25	0.8440	0.8205	3
x + x_hat	z + x_hat	0.0181623932	0.5	0.8440	0.8259	3
x + x_hat	z + skips	0.0085470085	0.5	0.8440	0.8355	3
x_hat + skips	x + skips	0.0181623932	0.25	0.8355	0.8173	3
x_hat + skips	skips	0.0096153846	0.75	0.8355	0.8259	3
x_hat + skips	z only	-0.0117521368	0.75	0.8355	0.8472	3
x_hat + skips	z + x	0.0149572649	0.5	0.8355	0.8205	3
x_hat + skips	z + x_hat	0.0096153846	0.75	0.8355	0.8259	3
x_hat + skips	z + skips	0.0000000000	1.0	0.8355	0.8355	3
x + skips	skips	-0.0085470085	1.0	0.8173	0.8259	3
x + skips	z only	-0.0299145299	0.5	0.8173	0.8472	3
x + skips	z + x	-0.0032051282	0.75	0.8173	0.8205	3
x + skips	z + x_hat	-0.0085470085	1.0	0.8173	0.8259	3
x + skips	z + skips	-0.0181623932	0.5	0.8173	0.8355	3
skips	z only	-0.0213675214	0.25	0.8259	0.8472	3
skips	z + x	0.0053418803	1.0	0.8259	0.8205	3
skips	z + x_hat	0.0000000000	1.0	0.8259	0.8259	3
skips	z + skips	-0.0096153846	0.5	0.8259	0.8355	3
z only	z + x	0.0267094017	0.25	0.8472	0.8205	3
z only	z + x_hat	0.0213675214	0.25	0.8472	0.8259	3
z only	z + skips	0.0117521368	0.5	0.8472	0.8355	3
z + x	z + x_hat	-0.0053418803	1.0	0.8205	0.8259	3
z + x	z + skips	-0.0149572649	0.25	0.8205	0.8355	3
z + x_hat	z + skips	-0.0096153846	0.75	0.8259	0.8355	3

Table 52: Pairwise permutation tests of decoder input ablations on seen subjects in the SSVEP task

config_1	config_2	mean_1	mean_2	p_value	statistic	n_seeds	p_value_corrected	significant
x + x_hat + skips	x + x_hat	0.8586	0.8516	0.5	0.00699	3	1	FALSE
x + x_hat + skips	x_hat + skips	0.8586	0.8570	0.8	0.00155	3	1	FALSE
x + x_hat + skips	x + skips	0.8586	0.8563	0.6	0.00233	3	1	FALSE
x + x_hat + skips	skips	0.8586	0.8403	0.2	0.01826	3	1	FALSE
x + x_hat + skips	z only	0.8586	0.8508	0.6	0.00777	3	1	FALSE
x + x_hat + skips	z + x	0.8586	0.8489	0.2	0.00971	3	1	FALSE
x + x_hat + skips	z + x_hat	0.8586	0.8516	0.1	0.00699	3	1	FALSE
x + x_hat + skips	z + skips	0.8586	0.8531	0.4	0.00544	3	1	FALSE
x + x_hat	x_hat + skips	0.8516	0.8570	0.5	-0.00544	3	1	FALSE
x + x_hat	x + skips	0.8516	0.8563	0.5	-0.00466	3	1	FALSE
x + x_hat	skips	0.8516	0.8403	0.4	0.01127	3	1	FALSE
x + x_hat	z only	0.8516	0.8508	1.0	0.00078	3	1	FALSE
x + x_hat	z + x	0.8516	0.8489	0.7	0.00272	3	1	FALSE
x + x_hat	z + x_hat	0.8516	0.8516	1.0	0.00000	3	1	FALSE
x + x_hat	z + skips	0.8516	0.8531	0.9	-0.00155	3	1	FALSE
x_hat + skips	x + skips	0.8570	0.8563	0.9	0.00078	3	1	FALSE
x_hat + skips	skips	0.8570	0.8403	0.3	0.01671	3	1	FALSE
x_hat + skips	z only	0.8570	0.8508	0.7	0.00622	3	1	FALSE
x_hat + skips	z + x	0.8570	0.8489	0.3	0.00816	3	1	FALSE
x_hat + skips	z + x_hat	0.8570	0.8516	0.1	0.00544	3	1	FALSE
x_hat + skips	z + skips	0.8570	0.8531	0.5	0.00389	3	1	FALSE
x + skips	skips	0.8563	0.8403	0.3	0.01593	3	1	FALSE
x + skips	z only	0.8563	0.8508	0.7	0.00544	3	1	FALSE
x + skips	z + x	0.8563	0.8489	0.3	0.00738	3	1	FALSE
x + skips	z + x_hat	0.8563	0.8516	0.3	0.00466	3	1	FALSE
x + skips	z + skips	0.8563	0.8531	0.6	0.00311	3	1	FALSE
skips	z only	0.8403	0.8508	0.5	-0.01049	3	1	FALSE
skips	z + x	0.8403	0.8489	0.5	-0.00855	3	1	FALSE
skips	z + x_hat	0.8403	0.8516	0.4	-0.01127	3	1	FALSE
skips	z + skips	0.8403	0.8531	0.3	-0.01282	3	1	FALSE
z only	z + x	0.8508	0.8489	0.8	0.00194	3	1	FALSE
z only	z + x_hat	0.8508	0.8516	1.0	-0.00078	3	1	FALSE
z only	z + skips	0.8508	0.8531	0.9	-0.00233	3	1	FALSE
z + x	z + x_hat	0.8489	0.8516	0.8	-0.00272	3	1	FALSE
z + x	z + skips	0.8489	0.8531	0.5	-0.00427	3	1	FALSE
z + x_hat	z + skips	0.8516	0.8531	0.9	-0.00155	3	1	FALSE

Table 53: Pairwise permutation tests of decoder input ablations on unseen subjects in the SSVEP task

config_1	config_2	mean_1	mean_2	p_value	statistic	n_seeds	p_value_corrected	significant
x + x_hat + skips	x + x_hat	0.8312	0.8440	0.6	-0.0128	3	1	FALSE
x + x_hat + skips	x_hat + skips	0.8312	0.8355	0.9	-0.0043	3	1	FALSE
x + x_hat + skips	x + skips	0.8312	0.8173	0.6	0.0139	3	1	FALSE
x + x_hat + skips	skips	0.8312	0.8259	0.9	0.0053	3	1	FALSE
x + x_hat + skips	z only	0.8312	0.8472	0.6	-0.0160	3	1	FALSE
x + x_hat + skips	z + x	0.8312	0.8205	0.5	0.0107	3	1	FALSE
x + x_hat + skips	z + x_hat	0.8312	0.8259	0.9	0.0053	3	1	FALSE
x + x_hat + skips	z + skips	0.8312	0.8355	0.9	-0.0043	3	1	FALSE
x + x_hat	x_hat + skips	0.8440	0.8355	0.8	0.0085	3	1	FALSE
x + x_hat	x + skips	0.8440	0.8173	0.5	0.0267	3	1	FALSE
x + x_hat	skips	0.8440	0.8259	0.6	0.0182	3	1	FALSE
x + x_hat	z only	0.8440	0.8472	0.8	-0.0032	3	1	FALSE
x + x_hat	z + x	0.8440	0.8205	0.5	0.0235	3	1	FALSE
x + x_hat	z + x_hat	0.8440	0.8259	0.7	0.0182	3	1	FALSE
x + x_hat	z + skips	0.8440	0.8355	0.6	0.0085	3	1	FALSE
x_hat + skips	x + skips	0.8355	0.8173	0.6	0.0182	3	1	FALSE
x_hat + skips	skips	0.8355	0.8259	0.8	0.0096	3	1	FALSE
x_hat + skips	z only	0.8355	0.8472	0.9	-0.0118	3	1	FALSE
x_hat + skips	z + x	0.8355	0.8205	0.6	0.0150	3	1	FALSE
x_hat + skips	z + x_hat	0.8355	0.8259	0.7	0.0096	3	1	FALSE
x_hat + skips	z + skips	0.8355	0.8355	1.0	0.0000	3	1	FALSE
x + skips	skips	0.8173	0.8259	0.9	-0.0085	3	1	FALSE
x + skips	z only	0.8173	0.8472	0.6	-0.0299	3	1	FALSE
x + skips	z + x	0.8173	0.8205	0.9	-0.0032	3	1	FALSE
x + skips	z + x_hat	0.8173	0.8259	0.8	-0.0085	3	1	FALSE
x + skips	z + skips	0.8173	0.8355	0.7	-0.0182	3	1	FALSE
skips	z only	0.8259	0.8472	0.5	-0.0214	3	1	FALSE
skips	z + x	0.8259	0.8205	1.0	0.0053	3	1	FALSE
skips	z + x_hat	0.8259	0.8259	1.0	0.0000	3	1	FALSE
skips	z + skips	0.8259	0.8355	0.6	-0.0096	3	1	FALSE
z only	z + x	0.8472	0.8205	0.5	0.0267	3	1	FALSE
z only	z + x_hat	0.8472	0.8259	0.6	0.0214	3	1	FALSE
z only	z + skips	0.8472	0.8355	0.6	0.0118	3	1	FALSE
z + x	z + x_hat	0.8205	0.8259	0.9	-0.0053	3	1	FALSE
z + x	z + skips	0.8205	0.8355	0.5	-0.0150	3	1	FALSE
z + x_hat	z + skips	0.8259	0.8355	0.8	-0.0096	3	1	FALSE

References

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

N. Kukhilava, T. Tsmindashvili, R. Kalandadze, A. Gupta, S. Katamadze, F. Brémond, L. M. Ferrari, P. Müller, and B. E. Wirth. Evaluation in eeg emotion recognition: State-of-the-art review and unified framework. *arXiv preprint arXiv:2505.18175*, 2025.

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.

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.

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

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.

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.

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

S. Nakagawa. A farewell to bonferroni: the problems of low statistical power and publication bias. *Behavioral ecology*, 15(6):1044–1045, 2004.

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.

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.

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.

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.

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.

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

F.-B. Vialatte and A. Cichocki. Split-test bonferroni correction for qeeg statistical maps. *Biological Cybernetics*, 98(4):295–303, 2008.

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

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

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.

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.