NEUROONLINE: BRIDGING PRETRAINING AND ONLINE ADAPTATION FOR EEG FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG foundation models have shown strong potential in learning generalized representations across subjects and tasks. However, most existing approaches follow a pretraining–static deployment paradigm, which suffers from two key limitations: (1) misalignment between pretraining objectives and downstream tasks, and (2) limited adaptability to distribution shifts in online settings. We propose Online Neural Adaptation (NeuroOnline), a unified framework that enables continuous adaptation in online scenarios. NeuroOnline integrates two complementary mechanisms: (1) multi-view consistency learning, which enforces cross-view alignment to promote consistent representations and improve robustness to distribution shifts; (2) context-aware representation modulation, which uses cross-attention between a learnable prompt and EEG representations to enable adaptive modulation under evolving data distributions. Together, these mechanisms unify representation alignment and dynamic adaptation. Experiments on multiple EEG benchmarks demonstrate that NeuroOnline consistently outperforms strong baselines in online settings and achieves robust performance under distribution shifts. Extensive experimental results further validate the effectiveness and robustness of the proposed framework. The code is available at `https://anonymous.4open.science/r/NeuroOnline-25FB`.

1 INTRODUCTION

Electroencephalography (EEG) is a key modality for brain-computer interfaces (BCIs), enabling characterization of neural activity for applications such as communication, control, cognitive monitoring, and neuromodulation (Abiri et al., 2019; Aggarwal & Chugh, 2022; Lotte et al., 2018; Värbu et al., 2022). Inspired by the success of large-scale pretraining in vision and natural language processing (Min et al., 2023; He et al., 2022; Chen et al., 2020; Devlin et al., 2019), recent studies have explored EEG foundation models to learn generalizable representations across subjects, tasks, and recording conditions (Kuruppu et al., 2026; Lai et al., 2025). These models exhibit strong cross-scenario generalization. This makes them promising candidates for general-purpose backbones in a wide range of downstream brain signal decoding tasks (Wu et al., 2025; Suhaimi et al., 2020; Boonyakitanont et al., 2020; Sharma et al., 2022; Amin et al., 2019; Roy et al., 2019; Biesmans et al., 2017; Aboalayon et al., 2016).

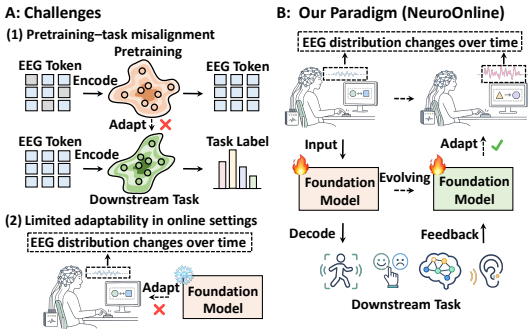

Figure 1: **Motivation.** (A) Challenges: pretraining–task misalignment and limited adaptability under distribution shifts in online EEG settings. (B) Our paradigm (NeuroOnline): continual adaptation of foundation models to evolving neural dynamics.

Despite recent progress, most EEG foundation models still adopt a pretraining–static deployment paradigm, which faces two fundamental challenges (Fig. 1A): **(1) Pretraining–task misalignment**.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Self-supervised objectives (e.g., masked reconstruction) inherently differ from downstream decoding goals, resulting in a mismatch between learned representations and task-specific data distributions (Liu et al., 2023; Wang et al., 2024b; Jiang et al., 2024). **(2) Limited adaptability under dynamic distribution shifts in online settings**. EEG signals are inherently non-stationary, with statistical properties that evolve over time, leading to distribution shifts during online deployment (Cecotti et al., 2025; Shen & Lin, 2019; Raza et al., 2019). In such settings, models must update their representations as new data arrive. However, without adaptive mechanisms, static models cannot adjust accordingly, further exacerbating the misalignment.

To address the above challenges, we propose a new learning paradigm, Online Neural Adaptation (NeuroOnline), that enables foundation models to continuously adapt to neural dynamics in online settings (Fig. 1B). NeuroOnline consists of two key mechanisms: **(1) Multi-view consistency learning**. We leverage observed online samples to generate multiple augmented views through stochastic transformations, such as temporal and frequency masking, and enforce consistency across them. This mechanism encourages consistent representations across augmented views, helping mitigate the distribution mismatch between pretraining and downstream tasks. **(2) Context-aware representation modulation.** We introduce a learnable prompt that interacts with incoming EEG representations through cross-attention to produce adaptive feature modulation, supporting continuous adaptation to evolving data distributions. These mechanisms correspond to representation alignment and dynamic adaptation, respectively, thereby bridging the gap between pretraining and online settings (Fig. 1). Our main contributions are summarized as follows:

- **Paradigm Innovation.** We propose Online Neural Adaptation (NeuroOnline), which shifts EEG foundation models from a static setting to continual adaptation, bridging the gap between pretraining and online settings.
- **Methodological Innovation.** We propose a multi-view consistency learning mechanism for pretraining–task alignment and a context-aware modulation mechanism for continuous adaptation under evolving data distributions.
- **Extensive Experiments.** We conduct extensive experiments across multiple foundation models and diverse EEG datasets, showing that NeuroOnline consistently outperforms strong baselines and demonstrates robust performance.

2 Related Work

2.1 EEG Foundation Models

Existing EEG foundation models aim to learn generalizable neural representations across subjects, tasks, and recording conditions (Kuruppu et al., 2026; Lai et al., 2025; Wu et al., 2025). This improves generalization in downstream brain signal decoding. Most approaches follow a paradigm of large-scale self-supervised pretraining with Transformer-based architectures to capture spatio-temporal dependencies. Representative methods include CBraMod (Wang et al., 2024b), LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024a), and BIOT (Yang et al., 2023), which differ in model design and training strategies but share the goal of learning scalable and transferable EEG representations. **However, these models are largely built upon a pretraining–static deployment paradigm. Their parameters remain fixed after deployment, with no mechanisms for continuous adaptation in online settings. As neural data distributions evolve over time, these models struggle to adapt effectively.**

2.2 Online Learning

Online learning has proven effective in computer vision and natural language processing, where continuously updating models with incoming data mitigates distribution shift and improves robustness in dynamic environments (Hoi et al., 2021; Chen et al., 2022; Huang et al., 2025; Shi et al., 2025; Wang et al., 2020). Inspired by this success, extending online learning to EEG representation learning provides a promising direction. **However, EEG signals exhibit stronger non-stationarity, greater inter-subject variability, and more complex temporal dynamics than image or text data (Jeon et al., 2023; Du et al., 2022), leading to more severe and harder-to-model distribution shifts in online settings. Consequently, developing dedicated online adaptation paradigms for**

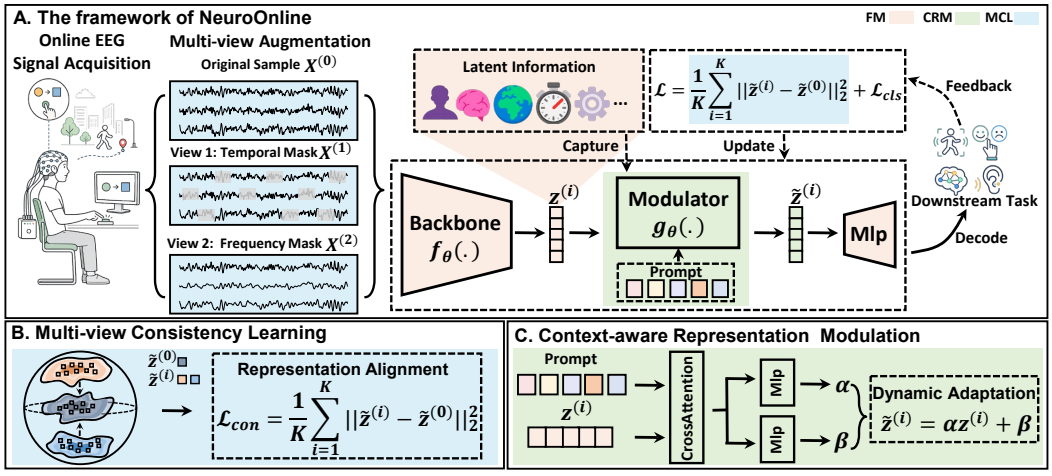

Figure 2: **Framework of NeuroOnline.** (A) Overall pipeline: given streaming EEG data, NeuroOnline constructs multi-view augmentations, encodes representations with a backbone model, and integrates multi-view consistency learning (MCL) and context-aware representation modulation (CRM) for continuous online adaptation. (B) Multi-view consistency learning: representations from augmented views are aligned to mitigate pretraining–task misalignment. (C) Context-aware representation modulation: a learnable prompt interacts with EEG representations through cross-attention to generate adaptive scaling and shifting parameters under evolving data distributions.

EEG foundation models remains an open challenge. To address this, we propose Online Neural Adaptation (NeuroOnline), which integrates the multi-view consistency learning mechanism with the context-aware representation modulation mechanism to enable continuous adaptation in online environments.

3 METHOD

3.1 PROBLEM FORMULATION AND OVERVIEW

Let $\{X_t\}_{t=1}^{T}$ denote a stream of sequentially arriving EEG segments, where each $X_t \in \mathbb{R}^{m \times N}$ represents a short temporal window with m frames and N channels. At step t, the model receives the current observation X_t and predicts the corresponding task label $y_t \in \mathbb{R}^d$, where d is the number of classes. Existing EEG foundation models typically follow a pretraining–static deployment paradigm, which suffers from two key limitations: (1) pretraining–task misalignment and (2) limited adaptability under distribution shifts in online settings. Our goal is to predict y_t accurately while continuously adapting the model to evolving data distributions.

We present NeuroOnline, a unified framework for online adaptation of EEG foundation models, as illustrated in Fig. 2. The framework comprises two key components: (i) multi-view consistency learning, which enforces cross-view alignment to learn consistent representations that better adapt to downstream EEG tasks, and (ii) context-aware representation modulation, which adaptively scales and shifts representations through prompt–representation interactions. Together, these components enable a unified online learning framework that mitigates pretraining–task misalignment and distribution shifts, thereby bridging the gap between pretraining and online adaptation.

3.2 MULTI-VIEW CONSISTENCY LEARNING (MCL)

To mitigate the pretraining–downstream task misalignment, we propose a multi-view consistency learning mechanism that promotes consistent representations across augmented views.

Multi-view Construction. We construct multiple augmented views from online samples observed up to time step t, ensuring a strictly causal setting. For each sample X, we define it as the anchor

view $X^{(0)}$ and generate K augmented views $\{X^{(i)}\}_{i=1}^{K}$ via stochastic transformations, including temporal masking and frequency masking, where $K = 2$. Specifically, temporal masking randomly removes contiguous segments along the time dimension to simulate local temporal corruption, while frequency masking randomly masks frequency bands to introduce spectral perturbations. These augmentations preserve semantic content while introducing complementary perturbations, enabling the model to learn consistent representations.

Consistency Learning. Given the constructed views, we enforce consistency in the modulated representation space. For each sample X, we compute the modulated representations for all views as:

$$\tilde{z}^{(i)} = g_\theta\big(f_\theta(X^{(i)})\big), \quad i = 0, \ldots, K, \tag{1}$$

where f_θ denotes the backbone model and g_θ the context-aware modulator, with $i = 0$ corresponding to the anchor view. We enforce consistency by aligning each augmented view with the anchor representation $\tilde{z}^{(0)}$ (Fig. 2B), with the loss defined as:

$$\mathcal{L}_{\text{cons}} = \frac{1}{K} \sum_{i=1}^{K} \left\| \tilde{z}^{(i)} - \tilde{z}^{(0)} \right\|_2^2. \tag{2}$$

By imposing consistency in the modulated representation space, the model learns view-consistent representations under evolving data distributions, helping mitigate the pretraining–task mismatch.

3.3 Context-aware Representation Modulation (CRM)

To accommodate evolving data distributions in online environments, we propose context-aware representation modulation. Here, context refers to latent distributional information induced through interactions between a learnable prompt and incoming EEG representations, rather than externally provided metadata.

Prompt–Representation Interaction. We maintain a learnable prompt $\mathcal{P}$ that interacts with each incoming EEG representation through cross-attention. The resulting interaction features condition the modulation parameters on the current input, allowing the representation to be adjusted as the online data distribution evolves.

Context-aware Modulation. Given the input representation $z = f_\theta(X)$, we modulate it using a prompt-conditioned function g_θ. Specifically, g_θ consists of a cross-attention module followed by two parallel MLP heads. The cross-attention module models the interaction between z and $\mathcal{P}$, while the two MLP heads generate the scaling and shifting parameters α and β, respectively. The modulated representation is given by:

$$\tilde{z} = g_\theta(z, \mathcal{P}) = \alpha \odot z + \beta, \tag{3}$$

where $\odot$ denotes element-wise multiplication. The same modulation is applied to all augmented views (see Fig. 2C for details). By conditioning the modulation parameters on prompt–representation interactions, CRM supports adaptive representation adjustment under evolving data distributions.

3.4 Training and Optimization

NeuroOnline first predicts each incoming sample and then performs incremental optimization once its label becomes available, following a strictly causal online protocol. The overall objective combines task-specific classification and multi-view consistency:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{cons}}, \tag{4}$$

where $\mathcal{L}_{\text{cls}}$ is binary cross-entropy with logits for binary tasks and cross-entropy for multi-class tasks, $\mathcal{L}_{\text{cons}}$ enforces consistency across augmented views, and λ controls their trade-off. The complete procedure is summarized in Algorithm 1 (see Appendix for details).

Table 1: Overview of downstream BCI tasks and datasets.

BCI Tasks	Datasets	Rate	# Channels	Duration	# Samples	Label
I. Motor Imagery Classification	BCIC-2A	250Hz	22	4s	5,088	4-class
	BCIC-2B	250Hz	3	4s	6,520	2-class
	SHU-MI	250Hz	32	4s	11,988	2-class
	PhysioNet-MI	160Hz	64	4s	9,837	4-class
II. Emotion Recognition	FACED	250Hz	32	10s	10,332	9-class
	SEED-V	1000Hz	62	1s	115,001	5-class
III. Mental Disorder Diagnosis	Mumtaz2016	256Hz	19	5s	7,143	2-class
IV. Mental Stress Detection	MentalArithmetic	500Hz	20	5s	1,707	2-class

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets: We evaluate the proposed method on eight widely used EEG decoding benchmarks across four downstream tasks (Table 1): (1) Motor Imagery Classification, including BCIC-2A (Brunner et al., 2008), BCIC-2B (Steyrl et al., 2016), SHU-MI (Ma et al., 2022), and PhysioNet-MI (Goldberger et al., 2000; Schalk et al., 2004); (2) Emotion Recognition, including FACED (Chen et al., 2023) and SEED-V (Liu et al., 2021); (3) Mental Disorder Diagnosis, including Mumtaz2016 (Wajid, 2016); (4) Mental Stress Detection, including MentalArithmetic (Zyma et al., 2019) (see Appendix B for details).

Baselines: Online continual adaptation for EEG foundation models remains underexplored, with no standardized baselines available. We therefore organize the competing methods into three representative groups. The first group comprises **offline methods**, which are trained offline and evaluated without online updates. We consider LoRA-based fine-tuning (Hu et al., 2022) and full fine-tuning, each with either a linear or an MLP head: **off-LoRA-LIN**, **off-LoRA-MLP**, **off-FT-LIN**, and **off-FT-MLP**. The second group consists of **online updating methods**. Since dedicated online adaptation approaches for EEG foundation models remain underexplored, we construct three representative baselines with different parameter-update strategies: (i) **on-FT-MLP**, which jointly updates the entire backbone and the MLP classifier; (ii) **on-LoRA-MLP**, which adapts the backbone using LoRA while jointly updating the MLP classifier; and (iii) **on-MLP**, which freezes the pretrained backbone and updates only the MLP classifier. The third group consists of **test-time training methods**, represented by **NeuroTTT** (Wang et al., 2025), an EEG-specific test-time training approach. These baselines enable a comprehensive evaluation of NeuroOnline (see Appendix C for details).

Backbone Models: We evaluate the proposed method on four representative EEG foundation models, including CBraMod (Wang et al., 2024b), LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024a), and BIOT (Yang et al., 2023) (see Appendix D for details).

Metrics: For binary classification, we report Balanced Accuracy (Bal.Acc.), AUC-PR, and AUROC, with AUROC used as the monitor metric. For multi-class classification, we report Balanced Accuracy (Bal.Acc.), Cohen's Kappa (Kappa), and Weighted F1 (F1-w), with Cohen's Kappa as the monitor metric (see Appendix E-G for details). All results are averaged over five random seeds and reported with standard deviations. $\Delta IMP \uparrow$ denotes the percentage relative improvement over the best competing method for each metric. ** and *** indicate statistical significance based on the two-sided Wilcoxon signed-rank test, corresponding to $p < 0.01$ and $p < 0.001$, respectively (Pratt, 1959).

Online Label Availability and Application Scenarios: NeuroOnline targets online EEG applications in which labels are available immediately or after a short delay. In cue-based motor-imagery neurorehabilitation, task cues specify the target movement for each trial, thereby providing trial-level labels for online adaptation, while decoded EEG activity can trigger contingent rehabilitation feedback or assistive actions. In interactive applications such as emotion recognition, labels may become available after prediction through user self-reports or explicit feedback, resulting in delayed but usable supervision. NeuroOnline first predicts each incoming sample with the currently deployed model and performs adaptation only after the corresponding label becomes available. This prediction-before-update protocol supports both immediate and delayed supervision while preserving causality and preventing target-label leakage.

Table 2: Performance comparison with baselines using CBraMod under online settings.

	BCIC-2A, 4-class			BCIC-2B, 2-class			SHU-MI, 2-class		
Methods	**Bal.Acc.**	**Kappa**	**F1-w**	**Bal.Acc.**	**AUC-PR**	**AUROC**	**Bal.Acc.**	**AUC-PR**	**AUROC**
off-LoRA-LIN	29.24 ± 2.31	5.65 ± 3.09	17.27 ± 3.40	50.35 ± 0.82	69.08 ± 1.49	69.56 ± 2.45	62.86 ± 0.25	69.21 ± 0.75	69.03 ± 0.54
off-LoRA-MLP	30.54 ± 1.26	7.38 ± 1.68	19.21 ± 1.48	60.01 ± 1.23	64.38 ± 1.70	65.18 ± 1.17	63.01 ± 0.69	68.55 ± 0.49	68.54 ± 0.61
off-FT-LIN	44.55 ± 4.11	26.06 ± 5.47	39.88 ± 5.47	66.14 ± 1.16	69.42 ± 1.22	71.87 ± 1.08	61.55 ± 1.09	66.93 ± 1.53	66.67 ± 1.69
off-FT-MLP	45.12 ± 4.75	26.83 ± 6.34	40.46 ± 7.17	65.59 ± 1.58	68.99 ± 1.13	71.44 ± 1.28	61.69 ± 0.75	68.55 ± 1.31	68.06 ± 1.81
on-FT-MLP	62.62 ± 1.52	50.16 ± 2.02	62.43 ± 1.63	64.86 ± 1.59	71.66 ± 1.16	71.62 ± 1.00	63.39 ± 0.98	68.53 ± 0.92	68.53 ± 1.08
on-LoRA-MLP	61.88 ± 1.27	49.17 ± 1.70	61.32 ± 1.41	63.05 ± 2.50	71.62 ± 1.79	70.97 ± 1.92	63.33 ± 0.93	65.68 ± 1.13	67.40 ± 0.67
on-MLP	60.71 ± 1.31	47.62 ± 1.74	60.24 ± 1.43	62.93 ± 1.31	71.32 ± 1.49	70.47 ± 1.32	63.50 ± 0.66	65.88 ± 0.98	67.59 ± 0.64
NeuroTTT	42.73 ± 1.83	23.63 ± 2.44	37.39 ± 3.57	66.19 ± 1.65	66.92 ± 2.14	70.78 ± 2.15	62.38 ± 0.86	68.59 ± 0.67	67.94 ± 1.25
NeuroOnline* **	**66.18** ± 2.11	**54.91** ± 2.81	**66.02** ± 2.11	**66.74** ± 0.61	**72.63** ± 1.55	**73.04** ± 1.29	**63.86** ± 1.13	**69.26** ± 0.81	**69.41** ± 0.81
$\Delta IMP \uparrow$	**+5.69%**	**+9.47%**	**+5.75%**	**+0.83%**	**+1.35%**	**+1.63%**	**+0.57%**	**+0.07%**	**+0.55%**
	PhysioNet-MI, 4-class			**FACED, 9-class**			**SEED-V, 5-class**		
Methods	**Bal.Acc.**	**Kappa**	**F1-w**	**Bal.Acc.**	**Kappa**	**F1-w**	**Bal.Acc.**	**Kappa**	**F1-w**
off-LoRA-LIN	58.20 ± 0.85	44.26 ± 1.14	57.99 ± 0.85	41.16 ± 0.89	33.55 ± 1.00	41.07 ± 0.95	37.88 ± 0.20	22.52 ± 0.24	38.15 ± 0.25
off-LoRA-MLP	58.03 ± 1.16	44.04 ± 1.55	58.05 ± 1.15	52.67 ± 0.47	46.29 ± 0.53	52.09 ± 0.38	40.07 ± 0.60	25.71 ± 0.86	41.05 ± 0.69
off-FT-LIN	60.58 ± 0.88	47.44 ± 1.17	60.36 ± 0.85	42.15 ± 0.75	34.56 ± 0.91	41.89 ± 0.96	37.40 ± 0.25	21.92 ± 0.27	37.98 ± 0.26
off-FT-MLP	60.46 ± 0.41	47.28 ± 0.55	60.36 ± 0.53	52.90 ± 0.39	46.67 ± 0.40	52.90 ± 0.33	39.35 ± 0.72	24.71 ± 1.03	40.19 ± 0.85
on-FT-MLP	53.89 ± 1.29	38.52 ± 1.73	53.89 ± 1.36	53.71 ± 0.29	47.48 ± 0.31	53.56 ± 0.30	44.87 ± 1.38	32.44 ± 1.87	46.36 ± 1.48
on-LoRA-MLP	57.05 ± 1.07	42.73 ± 1.43	57.07 ± 1.07	52.63 ± 0.45	46.21 ± 0.52	52.37 ± 0.45	41.83 ± 1.94	28.76 ± 2.64	43.47 ± 2.10
on-MLP	56.56 ± 1.09	42.08 ± 1.45	56.66 ± 1.07	52.42 ± 0.53	45.98 ± 0.61	52.16 ± 0.54	45.20 ± 1.83	33.17 ± 2.33	46.89 ± 1.86
NeuroTTT	60.73 ± 0.75	47.63 ± 1.01	60.47 ± 0.83	51.95 ± 3.65	46.09 ± 3.93	53.60 ± 2.57	39.28 ± 1.08	24.71 ± 1.46	40.13 ± 1.26
NeuroOnline* **	**61.74** ± 0.86	**48.99** ± 1.15	**61.76** ± 0.85	**54.03** ± 1.65	**47.78** ± 1.81	**53.76** ± 1.55	**59.37** ± 0.56	**50.35** ± 0.67	**60.56** ± 0.56
$\Delta IMP \uparrow$	**+1.66%**	**+2.86%**	**+2.13%**	**+0.60%**	**+0.63%**	**+0.30%**	**+31.35%**	**+51.79%**	**+29.15%**
	MentalArithmetic, 2-class			**Mumtaz2016, 2-class**			**Average**		
Methods	**Bal.Acc.**	**AUC-PR**	**AUROC**	**Bal.Acc.**	**AUC-PR**	**AUROC**	**Bal.Acc.**	**Kappa**	**AUROC**
off-LoRA-LIN	63.26 ± 3.52	45.50 ± 6.94	72.55 ± 3.21	88.02 ± 3.16	98.02 ± 0.35	97.83 ± 0.38	53.87 ± 1.50	26.50 ± 1.37	77.24 ± 1.65
off-LoRA-MLP	58.40 ± 5.14	42.12 ± 2.37	70.71 ± 5.74	88.12 ± 1.46	97.51 ± 0.23	97.23 ± 0.25	56.36 ± 1.50	30.86 ± 1.16	75.42 ± 1.94
off-FT-LIN	60.56 ± 5.97	49.74 ± 5.93	76.28 ± 3.12	88.22 ± 1.20	97.97 ± 0.73	97.73 ± 0.94	57.64 ± 1.93	32.50 ± 1.96	78.14 ± 1.71
off-FT-MLP	59.65 ± 2.66	48.67 ± 8.33	72.46 ± 5.91	88.64 ± 1.12	97.69 ± 0.97	97.53 ± 1.11	59.18 ± 1.55	36.37 ± 2.08	77.37 ± 2.53
on-FT-MLP	60.76 ± 3.28	48.23 ± 8.50	67.54 ± 4.34	98.70 ± 0.59	99.53 ± 0.10	99.26 ± 0.22	62.85 ± 1.37	42.15 ± 1.48	76.74 ± 1.66
on-LoRA-MLP	61.94 ± 2.27	52.13 ± 8.96	73.85 ± 4.37	98.69 ± 0.61	99.65 ± 0.31	99.44 ± 0.54	62.55 ± 1.38	41.72 ± 1.57	77.92 ± 1.88
on-MLP	62.22 ± 1.26	50.82 ± 7.83	72.50 ± 3.46	98.33 ± 0.56	99.66 ± 0.34	99.50 ± 0.58	62.73 ± 1.07	42.21 ± 1.53	77.52 ± 1.50
NeuroTTT	61.11 ± 4.12	49.71 ± 5.89	74.00 ± 5.02	88.86 ± 0.48	97.89 ± 0.54	97.74 ± 0.61	59.15 ± 1.80	35.52 ± 2.21	77.62 ± 2.26
NeuroOnline* **	**63.96** ± 2.77	**59.38** ± 6.69	**76.38** ± 5.38	**99.02** ± 0.22	**99.75** ± 0.13	**99.67** ± 0.23	**66.86** ± 1.24	**50.51** ± 1.61	**79.63** ± 1.93
$\Delta IMP \uparrow$	**+1.11%**	**+13.91%**	**+0.13%**	**+0.32%**	**+0.09%**	**+0.17%**	**+6.38%**	**+19.66%**	**+1.91%**

4.2 Performance Comparison with Baselines

Setup. We evaluate NeuroOnline on multiple datasets using CBraMod as the primary backbone, with all methods following the same data splits and online evaluation protocol. **Results.** As shown in Table 2, NeuroOnline achieves the best overall performance across tasks, datasets, and evaluation metrics. It outperforms the competing baselines in most evaluated settings, with statistically significant improvements ($p < 0.001$). The gains are particularly pronounced on challenging multi-class tasks. **Analysis.** Two main observations can be drawn: (i) Online updating generally outperforms static offline deployment, highlighting the importance of adapting EEG foundation models to distribution shifts. (ii) NeuroOnline achieves stronger and more consistent performance than the competing online approaches by jointly considering representation consistency and dynamic adaptation. Specifically, multi-view consistency learning encourages alignment across temporally and spectrally perturbed views, while context-aware representation modulation adaptively modulates representations through prompt–representation interactions. Their complementary effects contribute to improved online adaptation across diverse EEG decoding tasks.

4.3 Generalization Across EEG Foundation Models.

Setup. To assess the applicability of NeuroOnline beyond CBraMod, we further integrate it with LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024a) and BIOT (Yang et al., 2023). **Results and Analysis.** As shown in Table 3, NeuroOnline achieves the best overall performance across the evaluated backbones, datasets, and metrics, with statistically significant improvements over competing baselines ($p < 0.01$). These consistent gains across diverse architectures and pretraining strategies demonstrate that NeuroOnline is broadly compatible with EEG foundation models rather than being specific to CBraMod.

4.4 Ablation Study

Setup. We construct eight variants covering three aspects of the framework. For multi-view consistency learning, we remove the temporal-masking view (**w/o Time View**), the frequency-masking

Table 3: Performance comparison across additional EEG foundation models (LaBraM, EEGPT, BIOT) under online settings.

LaBraM

Methods	BCIC-2A, 4-class			FACED, 9-class			SEED-V, 5-class		
	Bal.Acc.	Kappa	F1-w	Bal.Acc.	Kappa	F1-w	Bal.Acc.	Kappa	F1-w
off-LoRA-LIN	46.37 ± 2.81	28.50 ± 3.74	44.22 ± 3.26	37.49 ± 0.71	29.32 ± 0.72	36.95 ± 0.52	39.90 ± 0.39	24.75 ± 0.57	40.17 ± 0.50
off-LoRA-MLP	46.09 ± 1.31	28.12 ± 1.75	44.06 ± 1.74	46.27 ± 1.13	39.34 ± 1.33	46.40 ± 1.12	41.34 ± 0.74	26.82 ± 1.01	41.80 ± 0.84
off-FT-LIN	47.40 ± 2.17	29.86 ± 2.89	45.25 ± 2.88	42.52 ± 0.88	35.04 ± 1.08	42.28 ± 0.99	39.29 ± 1.16	24.15 ± 1.51	39.64 ± 1.39
off-FT-MLP	51.13 ± 1.47	34.84 ± 1.96	49.75 ± 1.83	53.64 ± 0.42	47.75 ± 0.49	54.21 ± 0.48	41.24 ± 1.02	26.79 ± 1.48	41.63 ± 1.36
on-FT-MLP	62.08 ± 2.33	49.44 ± 3.11	61.97 ± 2.36	54.95 ± 0.30	48.97 ± 0.32	55.04 ± 0.32	41.27 ± 0.50	27.95 ± 0.71	42.70 ± 0.55
on-LoRA-MLP	64.60 ± 2.40	52.80 ± 3.20	64.44 ± 2.46	55.70 ± 0.68	49.79 ± 0.76	55.69 ± 0.65	41.82 ± 1.08	28.81 ± 1.41	43.41 ± 1.11
on-MLP	63.21 ± 2.82	50.95 ± 3.75	63.03 ± 2.85	55.19 ± 0.52	49.25 ± 0.57	55.21 ± 0.53	43.31 ± 0.93	30.67 ± 1.15	44.87 ± 0.90
NeuroTTT	51.01 ± 1.35	34.68 ± 1.80	49.44 ± 1.69	54.94 ± 1.14	49.10 ± 1.32	55.37 ± 1.21	39.66 ± 2.65	24.73 ± 3.70	39.77 ± 3.58
NeuroOnline**	**66.18** ± 0.85	**54.91** ± 1.13	**66.03** ± 0.86	**56.06** ± 0.82	**50.10** ± 0.90	**55.92** ± 0.79	**53.09** ± 0.58	**42.49** ± 0.68	**54.29** ± 0.55
$\Delta IMP \uparrow$	**+2.45%**	**+4.00%**	**+2.47%**	**+0.65%**	**+0.62%**	**+0.41%**	**+22.58%**	**+38.54%**	**+20.99%**

EEGPT

Methods	BCIC-2A, 4-class			FACED, 9-class			Mumtaz2016, 2-class		
	Bal.Acc.	Kappa	F1-w	Bal.Acc.	Kappa	F1-w	Bal.Acc.	AUC-PR	AUROC
off-LoRA-LIN	37.20 ± 2.10	16.27 ± 2.80	32.12 ± 3.62	28.39 ± 0.65	19.42 ± 0.69	28.12 ± 0.73	87.91 ± 3.12	96.62 ± 0.92	96.74 ± 0.83
off-LoRA-MLP	37.31 ± 1.02	16.41 ± 1.36	30.71 ± 1.74	33.60 ± 1.06	25.20 ± 1.20	33.51 ± 1.14	84.06 ± 3.92	95.32 ± 1.54	95.51 ± 1.42
off-FT-LIN	37.07 ± 2.19	16.09 ± 2.92	32.47 ± 3.07	33.70 ± 1.01	25.34 ± 1.11	33.58 ± 1.14	89.65 ± 1.79	96.41 ± 0.76	96.65 ± 0.54
off-FT-MLP	37.88 ± 2.15	17.18 ± 2.87	33.35 ± 2.62	39.60 ± 2.25	32.08 ± 2.45	39.96 ± 2.17	89.03 ± 2.06	96.55 ± 0.97	96.84 ± 0.54
on-FT-MLP	50.99 ± 1.35	34.65 ± 1.81	50.61 ± 1.52	43.08 ± 1.61	35.70 ± 1.78	43.04 ± 1.53	96.20 ± 1.50	99.35 ± 0.18	99.07 ± 0.22
on-LoRA-MLP	51.49 ± 1.20	35.32 ± 1.59	51.06 ± 1.41	42.03 ± 2.01	34.57 ± 2.22	42.10 ± 1.85	94.18 ± 1.70	99.23 ± 0.15	99.18 ± 0.14
on-MLP	51.01 ± 1.78	34.68 ± 2.37	50.54 ± 1.84	42.01 ± 2.03	34.54 ± 2.26	42.06 ± 1.92	96.27 ± 1.63	99.46 ± 0.39	99.25 ± 0.62
NeuroTTT	36.46 ± 2.31	15.28 ± 3.07	30.76 ± 4.04	38.35 ± 1.61	30.64 ± 1.81	38.57 ± 1.58	87.40 ± 3.74	97.07 ± 1.19	97.19 ± 1.11
NeuroOnline**	**53.35** ± 2.06	**37.80** ± 2.74	**52.46** ± 2.09	**49.28** ± 0.70	**42.59** ± 0.76	**49.39** ± 0.64	**97.31** ± 1.03	**99.57** ± 0.09	**99.36** ± 0.17
$\Delta IMP \uparrow$	**+3.61%**	**+7.02%**	**+2.74%**	**+14.39%**	**+19.30%**	**+14.75%**	**+1.08%**	**+0.11%**	**+0.11%**

BIOT

Methods	BCIC-2A, 4-class			Physionet-MI, 4-class			SEED-V, 5-class		
	Bal.Acc.	Kappa	F1-w	Bal.Acc.	Kappa	F1-w	Bal.Acc.	Kappa	F1-w
off-LoRA-LIN	30.24 ± 5.15	6.99 ± 6.87	19.46 ± 8.19	38.56 ± 3.23	18.07 ± 4.30	37.79 ± 3.44	34.13 ± 0.62	17.79 ± 0.73	34.74 ± 0.70
off-LoRA-MLP	34.11 ± 2.40	12.15 ± 3.20	26.72 ± 4.99	36.88 ± 3.04	15.83 ± 4.05	36.28 ± 3.16	34.39 ± 1.41	17.91 ± 1.64	34.83 ± 1.36
off-FT-LIN	30.05 ± 4.95	6.74 ± 6.60	18.98 ± 8.11	45.38 ± 3.41	27.17 ± 4.55	45.25 ± 3.43	35.47 ± 1.90	19.40 ± 2.29	35.98 ± 2.07
off-FT-MLP	29.43 ± 3.69	5.90 ± 4.92	19.26 ± 7.36	43.79 ± 4.06	25.04 ± 5.41	43.60 ± 4.02	36.17 ± 1.47	20.27 ± 1.74	36.80 ± 1.50
on-FT-MLP	55.45 ± 2.39	40.60 ± 3.19	55.38 ± 3.29	42.82 ± 2.10	23.77 ± 2.81	42.61 ± 2.08	38.28 ± 1.02	24.30 ± 1.26	39.76 ± 0.99
on-LoRA-MLP	51.60 ± 7.25	35.46 ± 9.67	51.51 ± 8.99	43.04 ± 3.95	24.05 ± 5.27	42.97 ± 3.95	34.83 ± 1.39	19.62 ± 1.72	36.05 ± 1.36
on-MLP	49.27 ± 6.40	32.36 ± 8.53	48.83 ± 8.08	42.72 ± 4.24	23.62 ± 5.65	42.69 ± 4.25	36.85 ± 0.96	22.01 ± 1.23	37.90 ± 0.95
NeuroTTT	29.64 ± 5.53	6.18 ± 7.37	18.41 ± 9.61	45.08 ± 2.56	26.75 ± 3.40	44.54 ± 2.36	35.65 ± 0.47	19.53 ± 0.53	36.04 ± 0.50
NeuroOnline**	**55.94** ± 6.72	**41.25** ± 8.97	**55.97** ± 7.73	**49.52** ± 3.30	**32.69** ± 4.41	**49.72** ± 3.33	**47.18** ± 0.77	**35.29** ± 1.03	**48.45** ± 0.83
$\Delta IMP \uparrow$	**+0.88%**	**+1.60%**	**+1.07%**	**+9.12%**	**+20.32%**	**+9.88%**	**+23.25%**	**+45.23%**	**+21.86%**

Table 4: Ablation experiment using CBraMod as the backbone.

Methods	BCIC-2A, 4-class			BCIC-2B, 2-class			MentalArithmetic, 2-class		
	Bal.Acc.	Kappa	F1-w	Bal.Acc.	AUC-PR	AUROC	Bal.Acc.	AUC-PR	AUROC
w/o Time View	63.66 ± 1.80	51.55 ± 2.40	63.49 ± 1.88	65.11 ± 0.30	69.97 ± 0.26	70.99 ± 0.32	59.93 ± 2.09	50.42 ± 1.33	74.37 ± 1.22
w/o Freq. View	63.85 ± 2.24	51.81 ± 2.98	63.64 ± 2.31	65.42 ± 0.34	70.18 ± 0.35	71.25 ± 0.34	59.93 ± 2.05	49.25 ± 1.24	73.77 ± 1.26
w/o Consistency	62.71 ± 1.04	50.28 ± 1.39	62.47 ± 1.17	65.50 ± 0.77	71.60 ± 0.59	71.48 ± 0.45	62.22 ± 1.97	48.82 ± 6.50	72.99 ± 3.49
w/o CRM	62.95 ± 1.03	50.60 ± 1.38	62.76 ± 1.07	65.73 ± 0.58	72.11 ± 0.67	72.20 ± 0.58	60.28 ± 1.96	48.19 ± 3.05	74.57 ± 0.97
CRM-MLP	64.57 ± 1.71	52.75 ± 2.28	64.38 ± 1.82	65.41 ± 0.66	70.48 ± 0.76	71.47 ± 0.65	60.21 ± 4.29	48.26 ± 8.23	72.10 ± 6.69
Prompt Addition	63.65 ± 2.16	51.53 ± 2.88	63.47 ± 2.26	65.62 ± 0.51	70.29 ± 1.01	71.48 ± 0.85	61.39 ± 4.35	46.12 ± 8.47	70.60 ± 6.29
w/o Prompt	64.57 ± 2.26	52.75 ± 3.01	64.38 ± 2.26	66.57 ± 0.99	72.55 ± 1.54	72.85 ± 1.35	59.44 ± 1.81	49.50 ± 3.61	72.99 ± 3.49
Frozen Random Prompt	65.28 ± 2.35	53.70 ± 3.14	65.12 ± 2.35	65.19 ± 0.75	70.06 ± 1.05	70.98 ± 0.97	59.44 ± 1.81	49.49 ± 3.60	74.07 ± 3.54
NeuroOnline	**66.18** ± 2.11	**54.91** ± 2.81	**66.02** ± 2.11	**66.74** ± 0.61	**72.63** ± 1.55	**73.04** ± 1.29	**63.96** ± 2.77	**59.38** ± 6.69	**76.38** ± 5.38

view (**w/o Freq. View**), or both views together with the consistency objective (**w/o Consistency**). For context-aware representation modulation, we remove the CRM module (**w/o CRM**), replace cross-attention with an MLP (**CRM-MLP**), or directly add the learnable prompt to the EEG representation without CRM (**Prompt Addition**). For prompt modeling, we remove the prompt (**w/o Prompt**) or use a randomly initialized and frozen prompt (**Frozen Random Prompt**). **Results and Analysis.** (i) As shown in Table 4, removing the temporal-masking view, the frequency-masking view, or both views together with the consistency objective generally leads to performance degradation, confirming the contribution of multi-view consistency learning. (ii) The CRM variants also underperform the full model in most settings, indicating that cross-attention-based prompt–representation interaction is more effective than removing CRM, replacing it with an MLP, or using direct prompt addition. (iii) Both removing the prompt and fixing it as a random vector reduce performance, highlighting the benefit of learnable prompt modeling. Overall, these results support the complementary roles of multi-view consistency learning, CRM, and the learnable prompt in NeuroOnline.

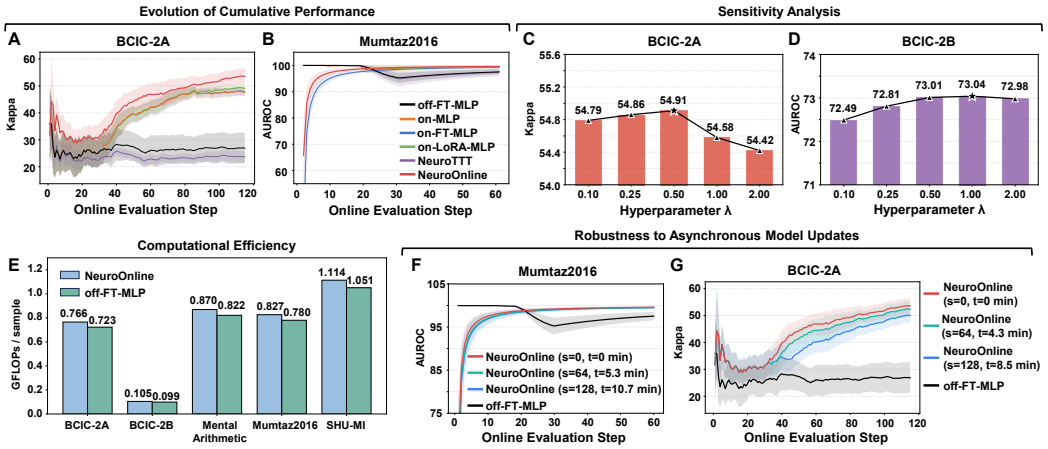

Figure 3: Comprehensive evaluation of NeuroOnline. (A–B) Cumulative performance evolution during online adaptation on BCIC-2A and Mumtaz2016. (C–D) Sensitivity analysis of the hyperparameter λ. (E) Computational efficiency comparison in terms of average FLOPs per sample across EEG datasets. (F–G) Robustness to Asynchronous Model Updates.

4.5 Evolution of Cumulative Performance.

Setup. To track performance throughout online adaptation, we compute cumulative metrics at periodic evaluation steps using all samples observed up to each step. We compare NeuroOnline with representative online updating and test-time training baselines on two representative datasets. **Results and Analysis.** As shown in Fig. 3(A-B), NeuroOnline maintains competitive cumulative performance and shows clearer advantages as online evaluation proceeds. On the two representative datasets, it achieves higher cumulative Kappa over most later evaluation steps in the multi-class task and maintains strong AUROC in the binary task. These trajectories indicate that the performance gains of NeuroOnline are sustained over the course of online evaluation rather than emerging only at the final step.

4.6 Sensitivity Analysis

To evaluate the sensitivity of the proposed method to key hyperparameters, we analyze the effect of the consistency loss weight λ on model performance. **Results:** As shown in Fig. 3(C-D), performance first improves and then degrades as λ increases. When λ is small, the consistency constraint is too weak to effectively enhance representation alignment. Conversely, overly large λ suppresses the classification objective, leading to performance decline. Notably, the model remains stable across a relatively wide range of λ. **Analysis:** These results demonstrate that the proposed method is robust to variations in the key hyperparameter and does not require careful tuning to achieve strong performance.

4.7 Computational Efficiency.

Setup. We evaluate the forward computational cost of NeuroOnline and off-FT-MLP, a standard full fine-tuning baseline, using average GFLOPs per sample. Both methods are measured with the same backbone and input configuration for a fair comparison. **Results and Analysis.** As shown in Fig. 3(E), NeuroOnline introduces slightly higher computational overhead due to its additional adaptation modules, while maintaining comparable overall costs with off-FT-MLP. These results demonstrate that NeuroOnline achieves effective online adaptation with only modest computational overhead.

4.8 Robustness to Asynchronous Model Updates.

Setup. To avoid blocking online prediction, NeuroOnline can be deployed with separate inference and adaptation processes. The inference process continuously handles incoming EEG segments using the current model, while the adaptation process updates a model replica in parallel and atomically activates the updated model upon completion. Since an update may not finish immediately, inference may temporarily rely on stale parameters. We therefore simulate synchronization delays of $s \in \{0, 64, 128\}$ EEG segments, where s denotes the number of subsequent EEG segments processed by the currently deployed model before the updated model is activated. The corresponding delays are approximately 0, 5.3, and 10.7 minutes on Mumtaz2016 and 0, 4.3, and 8.5 minutes on BCIC-2A, respectively. The case $s = 0$ corresponds to immediate synchronization. **Results and Analysis.** As shown in Fig. 3(F-G), NeuroOnline maintains comparable performance for delays of up to 128 EEG segments, with only limited degradation across the evaluated datasets. This indicates robustness to moderate parameter staleness. By decoupling inference from adaptation, the asynchronous design avoids blocking prediction while introducing only a bounded delay in deploying updated parameters, supporting practical deployment in latency-sensitive EEG applications.

5 Conclusion

In this work, we propose NeuroOnline, a unified framework for continuous adaptation of EEG foundation models in online environments. It integrates multi-view consistency learning and context-aware representation modulation to jointly model representation alignment and dynamic adaptation. Extensive experiments across multiple EEG benchmarks demonstrate that NeuroOnline consistently outperforms strong baselines in online settings. Extensive analyses further confirm its effectiveness and robustness. These results highlight the importance of jointly addressing representation alignment and dynamic adaptation for EEG foundation models in non-stationary environments. **Limitations and Future Work:** NeuroOnline currently relies on labels from previously observed online samples for adaptation and has mainly been evaluated in two representative application settings: motor-imagery-based neurorehabilitation and interactive emotion-recognition systems. Future work will investigate label-efficient and label-free adaptation and extend NeuroOnline to a broader range of EEG applications.

AI use statement

We did not use generative AI tools to generate scientific claims, conduct experiments, or analyze experimental results. The authors are responsible for the final content of this work.

Ethics statement

This work uses publicly available EEG datasets and does not involve new human subject experiments or data collection.

Reproducibility statement

To facilitate reproducibility, we provide detailed descriptions of the proposed framework, implementation details, and dataset preprocessing procedures. The anonymous source code repository is available at: `https://anonymous.4open.science/r/NeuroOnline-25FB`.

References

Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of eeg-based brain–computer interface paradigms. *Journal of neural engineering*, 16(1): 011001, 2019.

Khald Ali I. Aboalayon, Miad Faezipour, Wafaa S. Almuhammadi, and Saeid Moslehpour. Sleep stage classification using eeg signal analysis: A comprehensive survey and new investigation. *Entropy*, 18(9), 2016. ISSN 1099-4300. doi: 10.3390/e18090272. URL `https://www.mdpi.com/1099-4300/18/9/272`.

Swati Aggarwal and Nupur Chugh. Review of machine learning techniques for eeg based brain computer interface. *Archives of Computational Methods in Engineering*, 29(5):3001–3020, 2022.

Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed Amine Mekhtiche, and M. Shamim Hossain. Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion. *Future Generation Computer Systems*, 101:542–554, 2019. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2019.06.027. URL https://www.sciencedirect.com/science/article/pii/S0167739X19306077.

Wouter Biesmans, Neetha Das, Tom Francart, and Alexander Bertrand. Auditory-inspired speech envelope extraction methods for improved eeg-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(5): 402–412, 2017. doi: 10.1109/TNSRE.2016.2571900.

Poomipat Boonyakitanont, Apiwat Lek-uthai, Krisnachai Chomtho, and Jitkomut Songsiri. A review of feature extraction and performance evaluation in epileptic seizure detection using eeg. *Biomedical Signal Processing and Control*, 57:101702, 2020. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2019.101702. URL https://www.sciencedirect.com/science/article/pii/S1746809419302836.

Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. Bci competition 2008–graz data set a. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 16(1-6):1, 2008.

Hubert Cecotti, Rashmi Mrugank Shah, Raksha Jagadish, and Toshihisa Tanaka. Non-stationarity in brain-computer interfaces: An analytical perspective. *arXiv preprint arXiv:2512.15941*, 2025.

Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 295–305, 2022. doi: 10.1109/CVPR52688.2022.00039.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PmLR, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186, 2019.

Yang Du, Yongling Xu, Xiaoan Wang, Li Liu, and Pengcheng Ma. Eeg temporal–spatial transformer for person identification. *Scientific Reports*, 12(1):14378, 2022.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15979–15988, 2022. doi: 10.1109/CVPR52688.2022.01553.

Steven C.H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, 2021. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2021.04.112. URL https://www.sciencedirect.com/science/article/pii/S0925231221006706.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

Zhenpeng Huang, Xinhao Li, Jiaqi Li, Jing Wang, Xiangyu Zeng, Cheng Liang, Tao Wu, Xi Chen, Liang Li, and Limin Wang. Online video understanding: Ovbench and videochat-online. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 3328–3338, 2025.

Eunjin Jeon, Wonjun Ko, Jee Seok Yoon, and Heung-Il Suk. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):739–749, 2023. doi: 10.1109/TNNLS.2021.3100583.

Wei-Bang Jiang, Liming Zhao, and Bao-liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 16405–16426, 2024.

Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, and Yogatheesan Varatharajah. Eeg foundation models: a critical review of current progress and future directions. *Journal of neural engineering*, 23(2):021001, 2026.

Junhong Lai, Jiyu Wei, Lin Yao, and Yueming Wang. A simple review of eeg foundation models: Datasets, advancements and future perspectives. *arXiv preprint arXiv:2504.20069*, 2025.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2):715–729, 2021.

Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):857–876, 2023. doi: 10.1109/TKDE.2021.3090866.

Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

Jun Ma, Banghua Yang, Wenzheng Qiu, Yunzhe Li, Shouwei Gao, and Xinxing Xia. A large eeg dataset for studying cross-session variability in motor imagery brain-computer interface. *Scientific Data*, 9(1):531, 2022.

Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Comput. Surv.*, 56(2), September 2023. ISSN 0360-0300. doi: 10.1145/3605943. URL https://doi.org/10.1145/3605943.

John W. Pratt. Remarks on zeros and ties in the wilcoxon signed rank procedures. *Journal of the American Statistical Association*, 54:655–667, 1959. URL https://api.semanticscholar.org/CorpusID:120225998.

Haider Raza, Dheeraj Rathee, Shang-Ming Zhou, Hubert Cecotti, and Girijesh Prasad. Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface. *Neurocomputing*, 343:154–166, 2019. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.04.087. URL https://www.sciencedirect.com/science/article/pii/S0925231219301560.

Subhrajit Roy, Isabell Kiral-Kornek, and Stefan Harrer. Chrononet: A deep recurrent neural network for abnormal eeg identification. In *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings*, pp. 47–56, Berlin, Heidelberg, 2019. Springer-Verlag. ISBN 978-3-030-21641-2. doi: 10.1007/978-3-030-21642-9_8. URL https://doi.org/10.1007/978-3-030-21642-9_8.

Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

Lakhan Dev Sharma, Vijay Kumar Bohat, Maria Habib, Ala' M. Al-Zoubi, Hossam Faris, and Ibrahim Aljarah. Evolutionary inspired approach for mental stress detection using eeg signal. *Expert Systems with Applications*, 197:116634, 2022. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2022.116634. URL https://www.sciencedirect.com/science/article/pii/S0957417422001233.

Yi-Wei Shen and Yuan-Pin Lin. Challenge for affective brain-computer interfaces: non-stationary spatio-spectral eeg oscillations of emotional responses. *Frontiers in human neuroscience*, 13:366, 2019.

Haizhou Shi, Zihao Xu, Hengyi Wang, Weiyi Qin, Wenyuan Wang, Yibin Wang, Zifeng Wang, Sayna Ebrahimi, and Hao Wang. Continual learning of large language models: A comprehensive survey. *ACM Comput. Surv.*, 58(5), November 2025. ISSN 0360-0300. doi: 10.1145/3735633. URL https://doi.org/10.1145/3735633.

David Steyrl, Reinhold Scherer, Josef Faller, and Gernot R Müller-Putz. Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier. *Biomedical Engineering/Biomedizinische Technik*, 61(1):77–86, 2016.

Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Computational Intelligence and Neuroscience*, 2020(1):8875426, 2020. doi: https://doi.org/10.1155/2020/8875426. URL https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8875426.

Kaido Värbu, Naveed Muhammad, and Yar Muhammad. Past, present, and future of eeg-based bci applications. *Sensors*, 22(9):3331, 2022.

Mumtaz Wajid. Mdd patients and healthy controls eeg data (new). 2016. URL https://api.semanticscholar.org/CorpusID:151858102.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

Guagnyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 39249–39280. Curran Associates, Inc., 2024a. doi: 10.52202/079017-1239. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/4540d267eeec4e5dbd9dae9448f0b739-Paper-Conference.pdf.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. *arXiv preprint arXiv:2412.07236*, 2024b.

Suli Wang, Yangshen Deng, Zhenghua Bao, Xinyu Zhan, and Yiqun Duan. Neurottt: Bridging pretraining-downstream task misalignment in eeg foundation models via test-time training. *arXiv preprint arXiv:2509.26301*, 2025.

Jiamin Wu, Zichen Ren, Junyu Wang, Pengyu Zhu, Yonghao Song, Mianxin Liu, Qihao Zheng, Lei Bai, Wanli Ouyang, and Chunfeng Song. Adabrain-bench: Benchmarking brain foundation models for brain-computer interface applications. *arXiv preprint arXiv:2507.09882*, 2025.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 78240–78260. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/f6b30f3e2dd9cb53bbf2024402d02295-Paper-Conference.pdf.

Igor Zyma, Sergii Tukaev, Ivan Seleznov, Ken Kiyono, Anton Popov, Mariia Chernykh, and Oleksii Shpenkov. Electroencephalograms during mental arithmetic task performance. *Data*, 4(1):14, 2019.

A ALGORITHM

The overall optimization procedure of NeuroOnline is summarized in Algorithm 1.

Algorithm 1 The Pipeline of NeuroOnline

(1) Online Training Phase:
Input: Observed samples.
Step 1: Multi-view construction
For each sample X, treat it as the anchor view $X^{(0)}$ and generate K augmented views $\{X^{(i)}\}_{i=1}^{K}$ via stochastic transformations, including temporal and frequency masking.
Step 2: Context-aware representation modulation
Compute the modulated representations for the anchor and augmented views via Eq.(3): $\tilde{z}^{(i)} = g_\theta(f_\theta(X^{(i)}), \mathcal{P}),\ i = 0, \ldots, K$.
Step 3: Multi-view consistency learning
Compute consistency loss $\mathcal{L}_{\text{cons}}$ via Eq.(2),
Step 4: Online optimization
Compute overall loss $\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{cons}}$ via Eq.(4),
Update model parameters by minimizing $\mathcal{L}$.
(2) Online Inference Phase:
Input: Current EEG sample X_t,
Predict label $\hat{y}_t$ with fixed model parameters.

B ADDITIONAL DETAILS ON DATASETS AND PREPROCESSING

To better reflect real-world online settings, preprocessing is performed on individual EEG segments (i.e., epochs or trials) rather than on entire continuous EEG recordings. For each dataset, following the CBraMod, we split the data into training, validation, and test sets based on its original protocol or data structure. The training set is used for offline training or fine-tuning of baseline models, the validation set for hyperparameter selection, and the test set is streamed sequentially to simulate a realistic online setting, where we evaluate the model's continual adaptation capability.

Motor Imagery Classification.

BCIC-2A is a motor imagery (MI) dataset containing EEG recordings from 9 subjects using 22 channels. It includes four classes: left hand, right hand, both feet, and tongue movements. The continuous EEG signals are segmented into 4-second segments (i.e., epochs), resulting in 5088 samples. Each segment is band-pass filtered between 0.3–50 Hz and resampled to 200 Hz to remove low- and high-frequency noise. We adopt a strict cross-subject evaluation protocol, where subjects 1–5 are used for training, 6–7 for validation, and 8–9 for testing.

BCIC-2B is a motor imagery dataset consisting of EEG recordings from 9 subjects using 3 channels, with two classes: left- and right-hand motor imagery. The EEG signals are segmented into 4-second segments (i.e., epochs), yielding 6520 samples. Each epoch is band-pass filtered between 0–38 Hz and resampled to 200 Hz. To reduce artifacts, we remove electrooculogram (EOG) channels and apply common average referencing (CAR) across channels. We use the same cross-subject protocol as above, with subjects 1–5 for training, 6–7 for validation, and 8–9 for testing.

SHU-MI is a motor imagery (MI) EEG dataset containing recordings from 25 subjects with 32 channels, covering two classes: left- and right-hand imagery. The EEG signals are segmented into 4-second segments (i.e., epochs), yielding 11,988 samples, and are resampled to 200 Hz. We adopt a strict cross-subject protocol, using subjects 1–15 for training, 16–20 for validation, and 21–25 for testing.

PhysioNet-MI is a motor imagery (MI) EEG dataset containing recordings from 109 subjects with 64 channels, covering four classes: left fist, right fist, both fists, and both feet movements. The EEG signals are segmented into 4-second segments (i.e., epochs), yielding 9,837 samples. We apply common average referencing (CAR), followed by a 0.3 Hz high-pass filter to remove low-frequency drift. A 60 Hz notch filter is used to suppress power-line noise, and the signals are then resampled

to 200 Hz. We adopt a strict cross-subject protocol, using subjects 1–70 for training, 71–89 for validation, and 90–109 for testing.

Emotion Recognition.

FACED is a large-scale fine-grained affective EEG dataset containing recordings from 123 subjects with 32 channels, covering nine emotion classes: amusement, inspiration, joy, tenderness, anger, fear, disgust, sadness, and neutral. The EEG signals are segmented into 10-second segments, yielding 10,332 samples, and are resampled to 200 Hz. We adopt a cross-subject protocol, using subjects 1–80 for training, 81–100 for validation, and 101–123 for testing.

SEED-V is an affective EEG dataset containing recordings from 16 subjects with 62 channels, covering five emotion classes: happiness, sadness, neutral, disgust, and fear. Each subject has three sessions, and each session consists of 15 trials. According to the official documentation, one session is corrupted and is therefore excluded from our experiments. During preprocessing, we first remove irrelevant channels (M1, M2, VEO, and HEO) and extract trials based on the provided onset and offset annotations. The EEG signals are then resampled to 200 Hz and band-pass filtered between 0.3–75 Hz to remove low-frequency drift and high-frequency noise. Each trial is further segmented into non-overlapping 1-second segments, resulting in 115,001 samples. Within each session, the 15 trials are evenly divided into three parts (5:5:5) for training, validation, and testing, respectively.

Mental Disorder Diagnosis.

Mumtaz2016 is a clinical EEG dataset designed for major depressive disorder (MDD) diagnosis, consisting of recordings from 34 MDD patients and 30 healthy normal controls (NC). The EEG signals were collected using a 19-channel acquisition system with a sampling frequency of 256 Hz. The dataset includes three recording scenarios: eyes-open, eyes-closed, and task conditions. Consistent with the preprocessing protocol of CBraMod, we retain only the resting-state EEG recordings under the eyes-open and eyes-closed conditions. The raw EEG signals are first processed with a 0.3–75 Hz band-pass filter to eliminate low- and high-frequency artifacts, followed by a 50 Hz notch filter to reduce power-line noise. Subsequently, the signals are downsampled to 200 Hz and divided into 7,143 independent 5-second EEG segments without overlap. For subject-independent evaluation, 21 MDD subjects and 20 NC subjects are allocated for training, 5 MDD subjects and 4 NC subjects are used for validation, and the remaining 8 MDD subjects and 6 NC subjects are reserved for testing.

Mental Stress Detection.

MentalArithmetic is an EEG dataset developed for investigating mental stress induced by cognitive workload. It contains EEG recordings from 36 participants collected before and during a mental arithmetic experiment, with the pre-task and task periods corresponding to the non-stress and stress states, respectively. The recordings are obtained using 20 EEG channels at a sampling rate of 500 Hz. To remove irrelevant frequency components, the signals are filtered within the range of 0.5–45 Hz. Following the preprocessing pipeline of CBraMod, the EEG data are resampled to 200 Hz and segmented into 1,707 non-overlapping trials, each with a duration of 5 seconds. We employ a subject-wise split strategy to ensure cross-subject generalization, where subjects 1–28 are used for model training, subjects 29–32 are assigned for validation, and subjects 33–36 are reserved for final testing.

C Additional Details on Baselines

(1) Offline baselines.

Offline baselines are fine-tuned on the training set and directly evaluated on the test set without any online updates. To ensure fair comparison, all methods share the same pretrained backbone and follow identical data splits and optimization settings. We consider two categories of fine-tuning strategies, i.e., parameter-efficient adaptation (LoRA) and full fine-tuning (FT), each with either a linear or an MLP classification head:

(i) **off-LoRA-LIN:** We inject LoRA modules into the backbone and optimize only the low-rank parameters together with a linear classification head, while keeping the original backbone weights frozen;

(ii) **off-LoRA-MLP:** We adopt the same LoRA setting as in (i) but replace the linear head with an MLP classifier. Following CBraMod, the backbone output is flattened and passed through two hidden layers, each followed by ELU activation and dropout, before a final linear layer for prediction;

(iii) **off-FT-LIN:** We use a single linear classification head and jointly optimize the backbone and the classifier, serving as a lightweight full fine-tuning baseline;

(iv) **off-FT-MLP:** We attach the same MLP classification head as in (ii) and fine-tune the entire model end-to-end.

For LoRA-based methods, we uniformly apply LoRA modules to all linear layers in the backbone (including attention projections and feed-forward layers, except the output projection), enabling efficient adaptation without significantly increasing the number of parameters.

All offline methods are fixed after training and directly used for inference at test time, serving as baselines without online adaptability.

(2) Online baselines.

To ensure a fair comparison, all online baselines are implemented using the same backbone, input configuration, data splits, and online evaluation protocol as NeuroOnline. All methods follow the same causal online setting, where each incoming sample is first predicted by the current model without access to its label, and the label is used only for subsequent model adaptation. The main difference among baselines lies in the parameter update strategy during online adaptation.

(i) **Online Full Fine-tuning (on-FT-MLP):** This baseline updates all trainable parameters during online adaptation. Specifically, both the backbone and the MLP classifier are optimized. It serves as a full-parameter adaptation baseline to evaluate the benefit of updating the entire EEG foundation model in online scenarios.

(ii) **Online LoRA Fine-tuning (on-LoRA-MLP):** To investigate parameter-efficient online adaptation, we implement a LoRA-based baseline. The original backbone parameters are frozen, while additional LoRA parameters are inserted into the backbone and optimized together with the MLP classifier during online updates. This baseline evaluates whether low-rank parameter adaptation can effectively adapt EEG foundation models in online scenarios.

(iii) **Online Classifier Updating (on-MLP):** This baseline adopts a lightweight adaptation strategy by freezing the backbone and updating only the MLP classifier during online optimization. It adapts the decision boundary while keeping the pretrained EEG representations unchanged, serving as a representative baseline for classifier-only online adaptation.

(3) Test-Time training baseline.

NeuroTTT: We adopt the TTT with Tent variant of NeuroTTT, an EEG-specific test-time adaptation method for EEG foundation models. It minimizes prediction entropy on unlabeled EEG inputs and updates only normalization parameters during inference, while freezing the remaining parameters. This baseline evaluates entropy-based calibration under streaming EEG distribution shifts.

D More Details for Backbone Models

CBraMod adopts a criss-cross transformer with parallel spatial–temporal attention and masked pre-training to learn scalable and generalizable EEG representations.

LaBraM learns generalizable EEG representations via vector-quantized tokenization and masked Transformer pre-training on large-scale unlabeled data.

EEGPT employs a mask-based dual self-supervised objective with spatio-temporal representation alignment to learn high-quality and transferable EEG representations from low-SNR signals.

BIOT addresses heterogeneous biosignal formats by tokenizing multi-channel signals into a unified representation and modeling them with transformers for cross-dataset learning.

Hyperparameters	Settings
Epochs	50
Batch size	64
Dropout	0.1
Optimizer	AdamW
Learning rate	1e-4
Weight decay	5e-2
Scheduler	CosineAnnealingLR
Cosine cycle epochs	50
Minimal learning rate	1e-6
Clipping gradient norm	1
Label smoothing (multi-class)	0.1

Table 5: Hyperparameter settings for offline baselines.

Hyperparameters	Settings
Epochs	3
Batch size	64
Dropout	0.1
Optimizer	AdamW
Learning rate	1e-4
Weight decay	5e-2
Label smoothing (multi-class)	0.1

Table 6: Hyperparameter settings for online baselines and our method.

E Details of Evaluation Metrics

To comprehensively evaluate model performance, following CBraMod, we adopt several widely used metrics for downstream tasks.

Balanced Accuracy. Balanced Accuracy computes the average recall across classes, thereby mitigating the impact of class imbalance. It is used for both binary and multi-class classification.

AUC-PR. The area under the precision–recall curve (AUC-PR) reflects the trade-off between precision and recall across different thresholds and is particularly suitable for imbalanced binary classification.

AUROC. The area under the receiver operating characteristic curve (AUROC) measures the model's ability to distinguish between classes across varying decision thresholds. It is commonly used for evaluating binary classification performance.

Cohen's Kappa. Cohen's Kappa measures the agreement between predicted labels and ground truth while accounting for chance agreement. It naturally extends to multi-class settings and provides a more robust evaluation than accuracy, particularly when class distributions are imbalanced.

Weighted F1. Weighted F1 computes the average F1-score across classes, where each class is weighted by its sample proportion, providing a more comprehensive evaluation in multi-class classification.

F The Details of hyperparameter settings

According to CBraMod, the hyperparameter configurations for the offline baselines are provided in Table 5, while those for the online baselines and our method are reported in Table 6.

G Implementation Details

All experiments are implemented in Python 3.10.20 using PyTorch (version 2.11.0, CUDA 12.6). Experiments are conducted on a NVIDIA GeForce RTX 4090 GPU (24 GB memory).