When and How to Adapt: Subject Shifts Detection and Prototype-Guided Correction for Online EEG Decoding

Shaoqi Zhang[1,2] , **Xiyuan Jin**[1,2] , **Xiaojun Ning**[1,2] , **Yilin Chen**[1] and **Jing Wang**[1,2,3,*]
[1]School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China
[2]Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence, Beijing, China
[3]Key Laboratory of Big Data and Artificial Intelligence in Transportation, Ministry of Education, Beijing, China
{zhangsq, xiyuanjin, ningxj, chenyilin, wj}@bjtu.edu.cn

Abstract

Online EEG decoding is pivotal for real-world Brain-Computer Interfaces (BCIs) but confronts significant challenges arising from continuous distribution shifts, including both inter- and intra-subject variations. Existing Unsupervised Continual Domain Adaptation (UCDA) methods typically rely on rigid hard boundaries such as subject switch labels or fixed batch sizes to trigger adaptation. Meanwhile, some online EEG decoding methods that utilize detected shifts as soft boundaries are ill-suited for unsupervised scenarios and lack effective distribution alignment strategies. To address these issues, we propose a novel Prototype-driven online EEG Decoding framework (PRED). PRED incorporates a Subject Shift Detector (SSD) based on policy stability to reliably identify latent domain shifts unsupervisedly, thereby constructing adaptive soft boundaries. Furthermore, we design a Prototype-guided Shift Correction (PSC) mechanism that leverages a multi-granular memory structure to guide distribution alignment while preserving semantic stability. Experiments on three public datasets confirm that PRED achieves superior plasticity and stability, and demonstrates significant futurity, i.e., the capability of forward transfer to a subject's future samples.

1 Introduction

Electroencephalography (EEG) is a cornerstone of Brain-Computer Interfaces (BCI) due to its non-invasiveness and high temporal resolution. Deep learning has significantly advanced EEG decoding in tasks such as motor imagery [Herman *et al.*, 2008; Amin *et al.*, 2019], emotion recognition [Jenke *et al.*, 2014; Wang *et al.*, 2025], and sleep staging [Diykh *et al.*, 2016; Jia *et al.*, 2020]. However, with the proliferation of wearable EEG devices and real-time Human-Computer Interaction (HCI) systems, the research focus is shifting from offline paradigms to online reasoning and updating to cope with the distribution shifts of data streams in the temporal dimension [Li *et al.*, 2025], making robust online EEG decoding pivotal for practical applications.

*Corresponding author

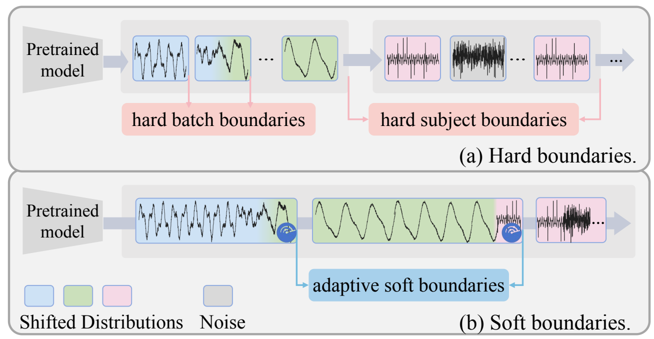

Figure 1: (a). Hard boundaries such as subject switch labels and fixed batch sizes; (b). Soft boundaries obtained by adaptive detection.

Online EEG data streams exhibit significant distribution shifts characterized by both intra-subject time-varying variations in long-term monitoring of wearable devices and inter-subject variations in multi-user deployment scenarios such as clinical assistance or human-computer interaction [Duan *et al.*, 2024a]. This compounded continuous distribution shifts coupled with the absence of real-time labeling makes online EEG decoding essentially an Unsupervised Continual Domain Adaptation (UCDA) problem.

Several existing methods [Taufique *et al.*, 2021; Taufique *et al.*, 2022; Chen *et al.*, 2025] have investigated the adaptation and generalization capabilities of models in UCDA scenarios. For instance, Zhou et al. [2025] proposed an Unsupervised Individual Continual Learning (UICL) framework, demonstrating favorable **plasticity** (i.e., the model's ability to adapt to new online subject distributions) and **stability** (i.e., the model's ability to generalize to the generalization set). However, as shown in Figure 1, relying exclusively on subject switch labels as hard boundaries to trigger adaptation completely overlooks intra-subject shifts, failing to update the model in time to improve online performance. While methods depending on fixed batch sizes as hard boundaries [Wang *et al.*, 2022; Cai *et al.*, 2021; Ye and Bors, 2024] struggle to balance adaptability and responsiveness: small batch updates are prone to noise-driven overfitting, while large batch updates often lead to adaptation lag, rendering the model ineffective against rapid or sudden distribution shifts. Therefore, autonomously capturing distribution shifts from streaming data to determine **when to adapt** (i.e., soft boundaries) remains a critical challenge in realistic online scenarios.

Furthermore, some online EEG decoding methods [Duan *et al.*, 2024a; Duan *et al.*, 2024b] attempt to detect soft boundaries by directly monitoring distribution changes in data representations, subsequently triggering adaptation. However, these methods still rely on target domain labels, making it impossible to accurately align the shifted target distribution with learned knowledge in the absence of ground truth guidance. Meanwhile, other approaches [Wang *et al.*, 2022; Zhou *et al.*, 2025] introduce only pseudo-labels, which act as unstable anchors, to guide the optimization process. When the model yields high-confidence incorrect predictions, forcing alignment with these noisy pseudo-labels causes errors to accumulate temporally, gradually dismantling the established semantic space and leading to severe negative transfer. Therefore, even if existing methods have partially alleviated the issue of adaptation timing, **how to adapt** in an effective and stable manner remains a key challenge.

To address the aforementioned challenges in real online scenarios, we propose a novel Prototype-driven online EEG Decoding framework (PRED). Operating in a plug-and-play manner, PRED enables the model to adaptively identify subject distribution shifts within online EEG data streams and perform shift correction, thereby achieving superior plasticity and stability. Furthermore, we explore a unique property of soft boundaries in online scenarios, namely **futurity** (i.e., the model learns a partial sample of a new subject and then improves its ability to generalize to future samples of that subject), which has been overlooked in current research. To this end, we design two novel modules: a Subject Shift Detector (SSD) based on policy stability [Schulman *et al.*, 2015] and a Prototype-guided Shift Correction (PSC) mechanism. Specifically, the SSD determines when to adapt by monitoring the relative cumulative deviation between the current classification policy and a reference classification policy. Upon detection of a significant shift, the PSC module is triggered. It leverages a multi-granular memory structure containing prototypes to perform experience replay and guide the correction of distribution shifts, enhancing the model's adaptability. Simultaneously, PSC maintains the semantic distribution anchored by prototypes to prevent performance deterioration caused by overfitting. The contributions of this paper are as follows:

- We design a Subject Shift Detector based on policy stability, which can reliably detect distribution shifts in online scenarios, thereby automatically determining the appropriate soft boundary for adaptation.
- We propose a Prototype-guided Shift Correction mechanism that not only enhances the model's rapid adaptation to new subject distributions but also improves the stability of model updates.
- We validate PRED on three public EEG decoding task datasets verified that PRED copes well with real online scenarios, achieving better plasticity and stability as well as exploring futurity in online settings.

2 Related Work

2.1 Domain Adaptation

To address the issue of cross-domain discrepancy, researchers have proposed numerous methods based on domain adaptation [She *et al.*, 2023b; She *et al.*, 2023a; Eldele *et al.*, 2021]. For instance, PR-PL [Zhou *et al.*, 2023] combines prototypical representation with pairwise learning to mitigate label noise and individual differences. A multi-source transfer learning framework [Li *et al.*, 2019] rapidly adapts models from existing subjects to new subjects utilizing a small amount of calibration data via source selection and style transfer mapping. Additionally, MSGDAN [Wang *et al.*, 2024a] combines dynamic graph alignment with a source selection mechanism to extract robust subject-invariant representations.

While these methods achieve effective adaptation for individual subjects, they neglect intra-subject variations. As a result, their direct application to real-world online scenarios with continuous subject distribution shifts often leads to cumulative errors.

2.2 Unsupervised Continual Domain Adaptation

Unsupervised Continual Domain Adaptation (UCDA) has garnered significant research attention to address the challenge of continuous domain shifts [Taufique *et al.*, 2021; Taufique *et al.*, 2023; Saporta *et al.*, 2022; Cho *et al.*, 2023]. For instance, UCL-GV [Taufique *et al.*, 2022] employs contrastive loss to better align the distribution discrepancy between the buffer and the continuous stream of target batches. BrainUICL [Zhou *et al.*, 2025] introduces a novel unsupervised individual continual learning paradigm, effectively balancing plasticity and stability [Mermillod *et al.*, 2013] during the adaptation process to new subjects. CoUDA [Chen *et al.*, 2025] leverages prototype contrastive learning and source-anchored knowledge distillation to achieve continual domain adaptation for industrial fault diagnosis. Furthermore, CoTTA [Wang *et al.*, 2022] reduces error accumulation via weight and augmentation averaging, while preventing catastrophic forgetting by stochastically restoring neurons to source weights.

The aforementioned UCDA methods typically trigger adaptations via rigid hard boundaries, such as subject switch labels or fixed batch sizes. Consequently, they suffer from adaptation lag and susceptibility to noise, making it difficult to effectively enhance online performance in real-world scenarios.

2.3 Online EEG Decoding

Recently, several studies have also focused on continuous EEG decoding to address cross-subject discrepancies [Duan *et al.*, 2023; Jin *et al.*, 2024; Li *et al.*, 2024a]. OMHGL [Pan *et al.*, 2023] fuses multimodal physiological signals via hypergraphs to capture high-order correlations and updates projections online to adapt to individual differences. To adaptively capture the soft boundary for triggering adaptation, MUDVI [Duan *et al.*, 2024a] employs kernel-based shift detection to identify changes in subject distribution and a memory buffer to mitigate forgetting. Similarly, AMBM [Duan *et al.*, 2024b] utilizes bayesian change point detection to trigger adaptation and bi-level mutual information maximization for knowledge retention.

Nevertheless, current online EEG decoding methods still rely on target domain labels and lack direct and effective distribution alignment strategies, rendering them ill-suited for realistic unsupervised online scenarios.

3 Preliminaries

Formally, let $\mathcal{D}_S = \{\mathcal{X}_S^i, \mathcal{Y}_S^i\}_{i=1}^{\mathcal{N}_S}$, $\mathcal{D}_T = \{\mathcal{X}_T^i\}_{i=1}^{\mathcal{N}_T}$, and $\mathcal{D}_G = \{\mathcal{X}_G^i, \mathcal{Y}_G^i\}_{i=1}^{\mathcal{N}_G}$ denote the labeled source, unlabeled online target, and labeled generalization domains, comprising $\mathcal{N}_S$, $\mathcal{N}_T$, and $\mathcal{N}_G$ subjects, respectively. In realistic online scenarios, we partition the data stream by detecting soft boundaries via the subject shift detector. We define the i-th incremental segment $\mathcal{X}_I^i$ as the samples collected between the $(i-1)$-th and i-th detected soft boundaries, forming the incremental domain $\mathcal{D}_I = \{\mathcal{X}_I^i\}_{i=1}^{\mathcal{N}_I}$ with $\mathcal{N}_I$ identified segments. The source pre-trained model M_0 is continuously updated on $\mathcal{D}_I$, yielding a sequence of models $M_0 \to \cdots \to M_{\mathcal{N}_I}$.

4 Methodology

Drawing inspiration from rehearsal-based continual learning methodologies [Lopez-Paz and Ranzato, 2017; Aljundi *et al.*, 2019; Rolnick *et al.*, 2019], this work leverages a multi-granular memory structure to implement a replay mechanism, thereby mitigating catastrophic forgetting. As illustrated in Figure 2, the model M_{i-1}, derived from the preceding incremental segment $\mathcal{X}_I^{i-1}$, continues to perform online EEG decoding. Upon the detection of a significant subject shift (i.e., a soft boundary), the prototype-guided shift correction mechanism is triggered to yield the adapted model M_i, which subsequently proceeds to the next iteration cycle.

Our proposed method functions as a plug-and-play online EEG decoding framework. The underlying model architecture comprises an encoder network $f(\cdot)$, a projection head $g(\cdot)$, and a classifier $\varphi(\cdot)$. An input sample x is mapped to a vector representation z via the encoder and the projection head, which is subsequently processed by the classifier to obtain the prediction vector $\hat{y}$:

$$z = g(f(x)),\ \hat{y} = \varphi(z). \tag{1}$$

4.1 Multi-granular Memory Structure

In neuroscience, the human memory system is organized into distinct functional hierarchies. Working memory acts as a temporary workspace for short-term retention [Baddeley, 2010], while the hippocampus facilitates consolidation by replaying experiences to integrate information [Berners-Lee *et al.*, 2022]. Ultimately, these memories evolve into stable and generalizable long-term knowledge within the neocortex [Squire *et al.*, 2015]. Mimicking this biological mechanism, we implement a Multi-granular Memory Structure (MMS): the Online buffer (B_o) for short-term data retention, the Replay buffer (B_r) for consolidating information via replay, and the Prototype buffer (B_p) for storing abstract prototypes as long-term memory. This architecture effectively combines rapid adaptation with long-term generalization to mitigate catastrophic forgetting.

The specific update process of the MMS is as follows: During the initialization phase prior to online decoding, we compute the mean vector representation of samples for each class in $\mathcal{D}_S$ to serve as prototypes:

$$P_c = \frac{1}{n_c} \sum_j z_j \cdot \mathbb{I}\{y_j = c)\}, \tag{2}$$

where n_c denotes the number of samples belonging to class c in $\mathcal{D}_S$, and $\mathbb{I}$ is the indicator function. This yields the prototype set $B_p = \{P_c\}_{c=1}^{N_c}$, containing N_c prototypes. Before adaptation is triggered, online data and their prediction vectors are stored in $B_o = \{x_o, \hat{y}_o\}_{o=1}^{N_o}$ in a First-In-First-Out (FIFO) manner, subject to the constraint that the capacity N_o remains less than the maximum capacity N_{max}. When adaptation is triggered, samples equal in quantity to B_o are class-balanced sampled from $\mathcal{D}_S$ to form $B_r = \{x_r, y_r\}_{r=1}^{N_o}$. Together with B_o and B_p, B_r is utilized for shift correction. Following adaptation, B_o and B_r are cleared, and B_p is updated using an Exponential Moving Average (EMA):

$$P_c^{\text{ema}} = \alpha \cdot P_c + (1 - \alpha) \cdot [\beta \cdot P_c^r + (1 - \beta) \cdot P_c^o], \tag{3}$$

where the replay prototype P_c^r is the mean vector representation of samples in class c from B_r, and the online prototype P_c^o is the mean vector representation of high-confidence samples in class c from B_o that satisfy $\max(\hat{y}_o) > \xi_1$. The parameters $\alpha \in [0, 1)$ and $\beta \in [0, 1)$ represent the smoothing coefficient for prototype updates and the weight of the replay prototypes, respectively. Given that the number of samples available for sampling in B_r is limited and B_o stores only a small amount of online data, the resulting storage cost is acceptable for online scenarios with continuous data streams.

4.2 Subject Shift Detector

Statistical shift detection is often prone to false positives due to task-irrelevant noise. Inspired by the trust-region constraint in TRPO [Schulman *et al.*, 2015], we design a Subject Shift Detector based on policy stability. By monitoring the relative cumulative deviation between the current and reference policies, this method identifies significant shifts. This task-oriented design circumvents the noise sensitivity inherent in pure data-level detection.

We regard the output probability distribution $q_t(y|x_t)$ of the classifier $\varphi(\cdot)$ for sample x_t at time step t as the current classification policy under state x_t:

$$q_t(y|x_t) = \text{softmax}(\hat{y}_t). \tag{4}$$

By calculating the cosine similarity between the sample feature z_t and the class prototypes in B_p, followed by normalization, we obtain the prototype similarity distribution $p_t(y|x_t)$ as the historical reference policy:

$$p_t(y|x_t) = \text{softmax}([\cos(z_t, P_1)/\tau, \ldots, \cos(z_t, P_{N_c})/\tau]), \tag{5}$$

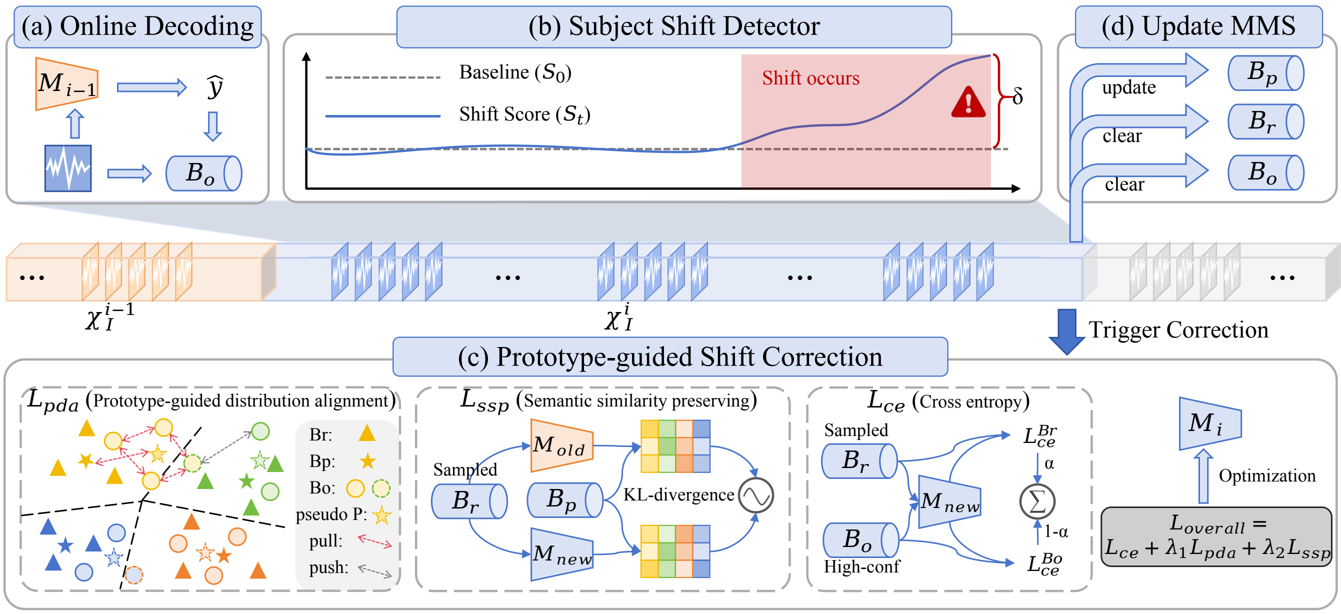

Figure 2: Workflow of the proposed PRED framework. During the pre-adaptation phase: (a) Online decoding: The model performs real-time online EEG decoding, storing the online samples and their prediction vectors into the online buffer within the Multi-granular Memory Structure (MMS); (b) Subject shift detector: The detector monitors distribution shifts and triggers shift correction when a specific threshold is reached. During the adaptation phase: (c) Prototype-guided shift correction: Guided by prototypes stored in the MMS, proximal samples in the new distribution are clustered and aligned toward the prototypes, while preserving the semantic distribution consistency of samples in the replay buffer before and after adaptation. In the post-adaptation phase: (d) Update MMS: The replay buffer and online buffer are cleared, and the prototype buffer is updated.

where τ denotes the temperature coefficient. The reference distribution is not a static obsolete policy, but rather a prior that updates gradually over time. This enables the detector to remain sensitive to short-term abrupt changes while accommodating acceptable long-term distribution drift. We employ the Kullback–Leibler (KL) divergence to compute the instantaneous deviation D_t between the current policy and the reference policy:

$$D_t = \mathrm{KL}(p_t(y|x_t) \parallel q_t(y|x_t)). \tag{6}$$

Given that the instantaneous deviation is susceptible to noise interference, we employ an EMA to derive the smoothed deviation S_t:

$$S_t = \gamma S_{t-1} + (1-\gamma) D_t, \tag{7}$$

where $\gamma \in [0, 1)$ represents the smoothing coefficient. To accommodate variations across different subject distributions, we estimate the baseline S_0 as the mean deviation calculated over the initial k samples of each detection phase:

$$S_0 = \frac{1}{k} \sum_{i=1}^{k} D_i. \tag{8}$$

Shift correction is triggered when $\frac{S_t}{S_0} - 1 > \delta$, indicating that the policy shift has exceeded the boundaries of the trust region. Here, the relative threshold δ ensures that the detector focuses on relative deviation rather than relying on an absolute scale.

4.3 Prototype-guided Shift Correction

Prototype-guided distribution alignment. While prototypes can serve as effective alignment anchors [De Lange and Tuytelaars, 2021; Li *et al.*, 2024b], online distribution shifts often cause samples to temporarily deviate towards mismatched prototypes in the feature space, making direct alignment to the nearest prototype prone to misalignment and noise amplification. We propose a prototype-guided strategy that clusters samples with similar prototype similarity distributions and identifies the target prototype via intra-cluster voting. By leveraging group structural consistency to suppress individual noise, this method ensures robust prototype-guided distribution alignment under distribution shifts.

First, we calculate the prototype similarity distribution p_i for sample $x_i \in B_o$ according to Eq. 5. This distribution not only characterizes the prediction tendency of the reference policy but also implicitly reflects the sample's relative position within the feature space. Consequently, samples sharing similar p_i are treated as potential intra-cluster neighbors, and a clustering-like effect is achieved via contrastive loss. Specifically, we construct the positive set $\mathcal{C}_i = \{x_j \in B_o \setminus \{x_i\} \mid \cos(p_i, p_j) > \xi_2\}$, where samples with similarity to p_i exceeding the threshold ξ_2 are grouped into the same cluster. The contrastive loss L_{con} is defined as:

$$L_{\text{con}} = \sum_{x_i \in B_o} \frac{-1}{|\mathcal{C}_i|} \sum_{x_j \in \mathcal{C}_i} \log \frac{\exp(\cos(p_i, p_j)/\tau)}{\sum_{x_k \in B_o \setminus \{x_i\}} \exp(\cos(p_i, p_k)/\tau)}. \tag{9}$$

Concurrently, we utilize the high-confidence pseudo-labels within the cluster for class determination, defining the predicted class $\tilde{y}_i$ of the cluster $\mathcal{C}_i$ via majority voting:

$$\tilde{y}_i = \arg\max_c \sum_{x_j \in \mathcal{C}_i} \mathbb{I}\{\arg\max(\hat{y}_j) = c\}. \tag{10}$$

Subsequently, we define the pseudo-prototype $\tilde{P}_i$ as the feature centroid of $\mathcal{C}_i$ and align it with its corresponding true prototype, yielding the alignment loss:

$$L_{\text{align}} = \sum_{x_i \in B_o} (1 - \cos(\tilde{P}_i, P_{\tilde{y}_i})). \tag{11}$$

By combining the components of the prototype-guided distribution alignment strategy, the total loss is formulated as:

$$L_{\text{pda}} = L_{\text{con}} + L_{\text{align}}. \tag{12}$$

Semantic similarity preserving. To prevent overfitting to outlier subject distributions during the model update process, which could disrupt the original discriminative structure and hinder subsequent recovery or lead to continuous performance degradation, it is imperative to preserve the semantic distribution throughout the update phase. Specifically, we denote the model undergoing the update process (comprising the encoder and projection head) and the model prior to triggering adaptation as $g_{\text{new}}(f_{\text{new}}(\cdot))$ and $g_{\text{old}}(f_{\text{old}}(\cdot))$, respectively. For each replay sample $x_r \in B_r$, we obtain its corresponding new and old vector representations, z_r^{new} and z_r^{old}:

$$z_r^{\text{new}} = g_{\text{new}}(f_{\text{new}}(x_r)),\ z_r^{\text{old}} = g_{\text{old}}(f_{\text{old}}(x_r)). \tag{13}$$

Subsequently, utilizing Eq. 5, we derive p_r^{new} and p_r^{old} to represent the new and old semantic distributions, respectively. We employ the KL divergence to enforce consistency in the semantic distribution during model updates, thereby formulating the semantic similarity preserving loss:

$$L_{\text{ssp}} = \frac{1}{N_r} \sum_{x_r \in B_r} \text{KL}(p_r^{\text{old}} \,\|\, p_r^{\text{new}}). \tag{14}$$

Overall loss function. In addition to the aforementioned losses, we incorporate a weighted cross-entropy loss targeting both the replay buffer B_r and the online buffer B_o:

$$\begin{aligned} L_{\text{ce}} &= \beta L_{\text{ce}}^{B_r} + (1-\beta) L_{\text{ce}}^{B_o} \\ &= -\beta \sum_{x_r, y_r \in B_r} y_r \log \varphi\big(g(f(x_r))\big) \\ &\quad - (1-\beta) \sum_{x_o, \hat{y}_o \in B_o} \hat{y}_o \log \varphi\big(g(f(x_o))\big). \end{aligned} \tag{15}$$

Here, the weighting parameter is set to be consistent with the weight β. Consequently, the overall objective function is formulated as:

$$L_{\text{overall}} = L_{\text{ce}} + \lambda_1 L_{\text{pda}} + \lambda_2 L_{\text{ssp}}, \tag{16}$$

where λ_1 and λ_2 denote the hyperparameters weighting the respective loss terms. By integrating prototype-guided distribution alignment and semantic similarity preserving, this overall loss function ultimately achieves robust correction of subject distribution shifts.

Dataset	# Subj.	# Chan.	Rate(Hz)	# Classes	Window (s)
Physionet-MI	105	64	160	4	4
FACED	123	32	128	9	10
ISRUC	100	8	100	5	30

Table 1: Summary of the public EEG datasets used for evaluation.

5 Experiments

5.1 Datasets

We evaluate PRED on three large-scale datasets (>100 subjects each) spanning Motor Imagery, Emotion Recognition, and Sleep Staging: Physionet-MI [Schalk *et al.*, 2004], FACED [Chen *et al.*, 2023], and ISRUC [Khalighi *et al.*, 2016]. As outlined in Table 1, this scale enables rigorous simulation of challenging long-term online decoding scenarios.

5.2 Experimental Settings and Implementation

Following the protocol established in BrainUICL [Zhou *et al.*, 2025], we partition each dataset into a source domain, an online target domain, and a generalization domain in a 3:5:2 ratio, with the initial model M_0 pre-trained on the source domain. The online target domain is utilized to simulate data streams for evaluating the model's plasticity and futurity, while the stability of each incremental model is assessed on the generalization domain after adapting to each incremental segment.

Specifically, during online EEG decoding, an incremental segment $\mathcal{X}_I^{i-1}$ is delineated when the model M_{i-2} detects a new subject distribution and triggers adaptation. Consequently, the subsequent incremental segment $\mathcal{X}_I^i$ is highly likely to contain future samples originating from the same subject distribution. Compared to the preceding model M_{i-2}, the model M_{i-1}, having adapted to $\mathcal{X}_I^{i-1}$, is expected to exhibit superior forward transfer on $\mathcal{X}_I^i$. We define this property as the model's futurity. In addition, the further performance enhancement of M_i on $\mathcal{X}_I^i$ subsequent to adaptation is termed plasticity, reflecting the capability for rapid adaptation to new subject distributions. To ensure a fair comparison, we utilize Global ACC (GACC) and Global Macro-F1 (GMF1) to measure plasticity and futurity, calculated based on the concatenation of predictions and labels across all incremental segments partitioned by either hard or soft boundaries (i.e., $\cup_{i=1}^{\mathcal{N}_I} \hat{\mathcal{Y}}_I^i$ and $\cup_{i=1}^{\mathcal{N}_I} \mathcal{Y}_I^i$). Furthermore, stability on the generalization domain is evaluated via Average Anytime Accuracy (AAA) [Caccia *et al.*, 2021] and Average Anytime Macro-F1 (AAF1). Here, AAA_i and AAF1_i denote the mean ACC and mean MF1, respectively, of the sequence of incremental models $\{M_0, M_1, \ldots, M_i\}$ on the generalization domain. The specific formulation is as follows:

$$\text{AAA}_i = \frac{1}{i+1} \sum_{j=0}^{i} \text{ACC}(\hat{\mathcal{Y}}_G^j,\ \mathcal{Y}_G), \tag{17}$$

and AAF1 is defined analogously using the MF1 score, where i indexes the i-th incremental segment; $\mathcal{Y}_G$ and $\hat{\mathcal{Y}}_G^j$ represent the ground-truth labels of the generalization domain and the

Dataset	Plasticity						Futurity						Stability					
	GACC			GMF1			GACC			GMF1			AAA			AAF1		
	M_0	M_i	$\Delta(\uparrow)$	M_0	M_i	$\Delta(\uparrow)$	M_{i-2}	M_{i-1}	$\Delta(\uparrow)$	M_{i-2}	M_{i-1}	$\Delta(\uparrow)$	M_0	$M_{\mathcal{N}_I}$	$\Delta(\uparrow)$	M_0	$M_{\mathcal{N}_I}$	$\Delta(\uparrow)$
Physionet-MI	46.41	49.43	**3.02**	45.51	49.55	**4.04**	48.62	49.33	**0.71**	48.74	49.43	**0.69**	49.89	51.75	**1.86**	48.87	51.46	**2.59**
FACED	34.91	42.02	**7.11**	35.11	42.53	**7.42**	40.02	41.21	**1.19**	40.58	41.74	**1.16**	38.59	42.81	**4.22**	38.74	43.07	**4.33**
ISRUC	69.76	74.32	**4.56**	64.86	72.37	**7.51**	71.84	73.17	**1.33**	69.80	71.12	**1.32**	73.22	74.92	**1.70**	69.23	73.01	**3.78**

Table 2: Overview of performance on three public EEG datasets across Plasticity, Futurity, and Stability metrics.

Dataset	Metric	UCDA							UDA	
		PRED-soft	PRED-hard	BrainUICL	CoUDA	DSS	COTTA	UCL-GV	PR-PL	TS-TCC
Physionet-MI	GACC	$\mathbf{49.43}_{\pm\mathbf{0.16}}$	$48.64_{\pm0.10}$	$48.56_{\pm0.28}$	$46.60_{\pm0.18}$	$46.29_{\pm0.25}$	$47.26_{\pm0.23}$	$45.87_{\pm0.20}$	$45.28_{\pm0.11}$	$43.93_{\pm1.05}$
	GMF1	$\mathbf{49.55}_{\pm\mathbf{0.20}}$	$48.71_{\pm0.19}$	$47.49_{\pm0.37}$	$46.46_{\pm0.17}$	$45.07_{\pm0.41}$	$45.66_{\pm0.28}$	$44.60_{\pm0.22}$	$42.13_{\pm0.11}$	$44.12_{\pm1.04}$
	AAA	$\mathbf{51.75}_{\pm\mathbf{0.35}}$	$50.35_{\pm0.29}$	$50.02_{\pm0.33}$	$50.45_{\pm0.04}$	$49.84_{\pm0.20}$	$50.01_{\pm0.01}$	$49.23_{\pm0.15}$	$47.42_{\pm0.04}$	$45.66_{\pm0.50}$
	AAF1	$\mathbf{51.46}_{\pm\mathbf{0.42}}$	$50.18_{\pm0.35}$	$49.69_{\pm0.46}$	$50.46_{\pm0.04}$	$48.44_{\pm0.34}$	$47.86_{\pm0.01}$	$47.92_{\pm0.26}$	$44.38_{\pm0.05}$	$44.27_{\pm0.60}$
FACED	GACC	$\mathbf{42.02}_{\pm\mathbf{0.24}}$	$41.18_{\pm0.36}$	$40.19_{\pm0.37}$	$39.18_{\pm0.20}$	$35.75_{\pm0.37}$	$37.74_{\pm0.18}$	$36.71_{\pm0.12}$	$35.84_{\pm0.17}$	$34.47_{\pm0.39}$
	GMF1	$\mathbf{42.53}_{\pm\mathbf{0.26}}$	$41.61_{\pm0.33}$	$39.67_{\pm0.36}$	$39.60_{\pm0.24}$	$35.69_{\pm0.36}$	$37.63_{\pm0.16}$	$36.66_{\pm0.08}$	$35.13_{\pm0.15}$	$34.44_{\pm0.40}$
	AAA	$\mathbf{42.81}_{\pm\mathbf{0.15}}$	$42.16_{\pm0.16}$	$40.79_{\pm0.17}$	$40.10_{\pm0.04}$	$37.91_{\pm0.13}$	$37.22_{\pm0.02}$	$38.90_{\pm0.12}$	$37.60_{\pm0.03}$	$36.90_{\pm0.31}$
	AAF1	$\mathbf{43.07}_{\pm\mathbf{0.15}}$	$42.42_{\pm0.17}$	$40.98_{\pm0.16}$	$40.33_{\pm0.04}$	$37.88_{\pm0.12}$	$37.05_{\pm0.02}$	$38.81_{\pm0.11}$	$37.66_{\pm0.04}$	$36.75_{\pm0.33}$
ISRUC	GACC	$\mathbf{74.32}_{\pm\mathbf{0.33}}$	$73.78_{\pm0.36}$	$73.97_{\pm0.61}$	$73.30_{\pm0.37}$	$70.73_{\pm0.41}$	$70.05_{\pm0.32}$	$70.91_{\pm0.24}$	$72.32_{\pm0.11}$	$65.48_{\pm1.50}$
	GMF1	$\mathbf{72.37}_{\pm\mathbf{0.34}}$	$71.74_{\pm0.33}$	$69.78_{\pm0.60}$	$67.47_{\pm0.45}$	$64.95_{\pm0.39}$	$64.20_{\pm0.64}$	$66.42_{\pm0.20}$	$66.48_{\pm0.10}$	$60.91_{\pm1.65}$
	AAA	$\mathbf{74.92}_{\pm\mathbf{0.36}}$	$74.46_{\pm0.35}$	$74.25_{\pm0.26}$	$73.83_{\pm0.50}$	$71.67_{\pm0.42}$	$73.17_{\pm0.34}$	$71.62_{\pm0.14}$	$71.47_{\pm0.03}$	$66.59_{\pm1.40}$
	AAF1	$\mathbf{73.01}_{\pm\mathbf{0.31}}$	$72.43_{\pm0.37}$	$70.37_{\pm0.43}$	$67.26_{\pm0.75}$	$66.22_{\pm0.48}$	$67.94_{\pm0.64}$	$67.13_{\pm0.18}$	$66.03_{\pm0.03}$	$61.75_{\pm1.77}$

Table 3: Comparison with baselines on Physionet-MI, FACED, and ISRUC datasets.

predictions generated by model M_j, respectively. Subsequent experiments are conducted over 5 independent runs with the subject input order in the online target domain randomly shuffled (without altering data partitions), and statistical results are reported.

PRED is implemented in PyTorch using the Adam optimizer. For adaptation, the learning rate is set to 5×10^{-6} with 10 epochs. The online memory capacity N_{max} is set to the average number of samples per subject in the dataset. The coefficients α, β, γ, and threshold ξ_1, ξ_2 are all set to 0.9. The Subject Shift Detector uses $k = 20$ and $\delta = 0.05$, while loss weights λ_1 and λ_2 are set to 0.1 and 0.5, respectively.

5.3 Overview Performance

As presented in Table 2, we evaluated the PRED framework across three datasets. Specifically, for the i-th incremental segment $\mathcal{X}_I^i$, we assessed performance using models at four distinct temporal stages: M_0, M_{i-2}, M_{i-1}, and M_i. Additionally, we evaluated the performance of the most recently updated model on the generalization domain following each adaptation step. Here, M_0 denotes the initial pre-trained model; M_{i-2} represents the model prior to any exposure to the new subject's distribution; M_{i-1} signifies the model that has learned from historical samples of the new subject distribution but has not yet encountered the future samples contained within $\mathcal{X}_I^i$; and M_i corresponds to the model fully adapted to the current incremental segment $\mathcal{X}_I^i$. Finally, $M_{\mathcal{N}_I}$ represents the final model resulting from continuous adaptation to the entire online data stream. Regarding plasticity, the updated M_i consistently exhibits performance improvements compared to M_{i-1}. This improvement is even more pronounced when compared to M_0. Regarding futurity, by learning from a limited number of samples from the new subject distribution, M_{i-1} achieves substantial gains relative to M_{i-2}. Regarding stability, unlike certain domain incremental models that merely mitigate forgetting rates, our approach further enhances generalization capabilities on the generalization domain through online decoding learning, and eventually significantly outperforms M_0 .

5.4 Comparison with Other Methods

Under the hard boundary setting, we implemented several UDA and UCDA baselines, including TS-TCC [Eldele *et al.*, 2021], PR-PL [Zhou *et al.*, 2023], UCL-GV [Taufique *et al.*, 2022], COTTA [Wang *et al.*, 2022], DSS [Wang *et al.*, 2024b], CoUDA [Chen *et al.*, 2025], and BrainUICL [Zhou *et al.*, 2025], while evaluating two variants of PRED: PRED-hard (using hard boundaries) and PRED-soft (using soft boundaries). As Futurity is not applicable under the hard boundary setting and the initial model remains consistent, we exclusively report plasticity (measured via GACC and GMF1 of M_i) and stability (measured via AAA and AAF1 of $M_{\mathcal{N}_I}$).

Table 3 shows PRED outperforms all baselines in both plasticity and stability. UDA baselines performed poorest due to their inability to handle continuous adaptation, while other UCDA methods fared better by accounting for continual domain shifts. PRED-hard outperforms baselines on most metrics in all three datasets, leveraging prototype-guided subject distribution alignment and semantic preservation to boost adaptability and generalization. PRED-soft further elevates performance, as the soft boundaries identified by the SSD facilitate timely adaptation to subject shifts. This mechanism avoids negative transfer from non-task noise, ultimately striking a superior balance between plasticity and stability.

Method	**Plasticity**						**Futurity**						**Stability**					
	GACC			GMF1			GACC			GMF1			AAA			AAF1		
	M_0	M_i	$\Delta(\uparrow)$	M_0	M_i	$\Delta(\uparrow)$	M_{i-2}	M_{i-1}	$\Delta(\uparrow)$	M_{i-2}	M_{i-1}	$\Delta(\uparrow)$	M_0	$M_{\mathcal{N}_I}$	$\Delta(\uparrow)$	M_0	$M_{\mathcal{N}_I}$	$\Delta(\uparrow)$
Base	69.76	71.87	2.11	64.86	69.83	4.97	71.11	71.51	0.40	68.61	69.25	0.64	73.22	73.79	0.57	69.23	71.58	2.35
Base+L_{pda}	69.76	73.37	3.61	64.86	71.48	6.62	71.36	72.06	0.70	69.27	70.16	0.89	73.22	74.51	1.29	69.23	72.63	3.40
Base+L_{ssp}	69.76	73.08	3.32	64.86	71.34	6.48	71.72	72.41	0.69	69.90	70.62	0.72	73.22	74.48	1.26	69.23	72.65	3.42
RSID	69.76	73.31	3.55	64.86	71.62	6.76	72.06	72.79	0.73	70.15	71.05	0.90	73.22	73.93	0.71	69.23	72.05	2.82
BOCPD	69.76	73.01	3.25	64.86	71.18	6.32	71.42	71.99	0.57	69.39	70.16	0.77	73.22	73.76	0.54	69.23	71.83	2.60
PRED	**69.76**	**74.32**	**4.56**	**64.86**	**72.37**	**7.51**	**71.84**	**73.17**	**1.33**	**69.80**	**71.12**	**1.32**	**73.22**	**74.92**	**1.70**	**69.23**	**73.01**	**3.78**

Table 4: Performance comparison with ablated methods.

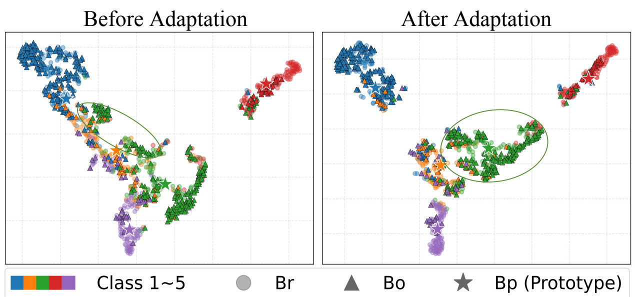

Figure 3: Visualization before and after adaptation on ISRUC.

5.5 Ablation Study

We conducted a two-fold ablation study on ISRUC, focusing on the contributions of the loss functions and the superiority of the subject shift detector (Table 4).

Ablation of loss terms. We evaluated variants including Base (L_{ce} only), Base + L_{pda}, and Base + L_{ssp}. Results indicate that L_{ssp} significantly enhances stability by constraining semantic consistency to prevent overfitting, which subsequently bolsters plasticity and futurity. Meanwhile, L_{pda} excels in plasticity by improving structural consistency and prototype alignment. The full PRED integrates both, creating a synergistic effect: L_{pda} ensures rapid adaptation to shifts, while L_{ssp} guarantees stable updates, collectively achieving optimal performance across all metrics.

Ablation of the subject shift detector. We compared our SSD against RSID [Duan *et al.*, 2024a] and BOCPD [Duan *et al.*, 2024b]. SSD outperforms both across all metrics. Unlike RSID and BOCPD, which rely on data-level detection and are thus susceptible to noise-induced false positives and subsequent negative transfer, SSD triggers only on significant policy deviations. By effectively filtering task-irrelevant statistical drifts, SSD ensures more reliable soft boundaries.

5.6 Visualization

Visualization of feature adaptation dynamics. We visualized feature distribution evolution on ISRUC via t-SNE (Figure 3). The left panel reveals significant shifts between online data (B_o) and established memory (B_r, B_p) upon SSD triggering (e.g., the green Class 3), validating the detector's sensitivity to subject shifts and the correctness of the soft boundaries. The right panel demonstrates that PSC successfully realigns shifted samples into clusters with clearer boundaries.

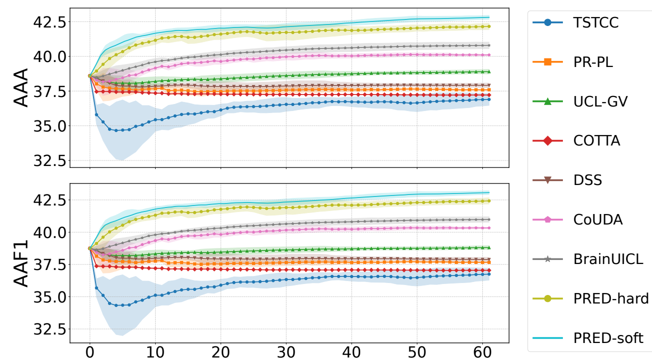

Figure 4: The AAA and AAF1 curves of PRED and baselines on FACED. These curves represent the performance of incremental models on the generalization domain, with shaded areas indicating the 95% confidence interval. Note that for PRED-soft, the curve is fitted to visualize the trend, as the specific timing of adaptation triggers varies across different input sequences.

This effective correction mechanism underpins PRED's superior plasticity and futurity.

Visualization of stability evolution. Figure 4 illustrates that the AAA and AAF1 curves for PRED exhibit a significant upward trend that gradually stabilizes, where PRED-soft displays better stability than PRED-hard. Enabled by the SSD to identify inter- and intra-subject shifts for precise soft boundary definition, PRED achieves effective adaptation and improved generalization. Compared to baseline methods, PRED exhibits superior stability.

6 Conclusion

We presented PRED, a novel prototype-driven framework for online reliable soft boundary detection and effective shift correction. By integrating the Subject Shift Detector (SSD) and Prototype-guided Shift Correction (PSC), PRED effectively solves the fundamental problem of when and how to adapt in unsupervised data streams. Specifically, SSD reliably identifies latent distribution shifts based on policy stability (when), while PSC leverages multi-granular memory structure to guide robust alignment (how). Extensive experiments across three public datasets demonstrate that PRED achieves a superior balance between plasticity and stability, and notably, exhibits significant futurity, paving the way for more realistic and autonomous BCI systems.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62576027).

References

[Aljundi *et al.*, 2019] Rahaf Aljundi, Min Lin, Baptiste Goujaud, and Yoshua Bengio. Gradient based sample selection for online continual learning. *Advances in neural information processing systems*, 32, 2019.

[Amin *et al.*, 2019] 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.

[Baddeley, 2010] Alan Baddeley. Working memory. *Current biology*, 20(4):R136–R140, 2010.

[Berners-Lee *et al.*, 2022] Alice Berners-Lee, Ting Feng, Delia Silva, Xiaojing Wu, Ellen R Ambrose, Brad E Pfeiffer, and David J Foster. Hippocampal replays appear after a single experience and incorporate greater detail with more experience. *Neuron*, 110(11):1829–1842, 2022.

[Caccia *et al.*, 2021] Lucas Caccia, Rahaf Aljundi, Nader Asadi, Tinne Tuytelaars, Joelle Pineau, and Eugene Belilovsky. New insights on reducing abrupt representation change in online continual learning. *arXiv preprint arXiv:2104.05025*, 2021.

[Cai *et al.*, 2021] Zhipeng Cai, Ozan Sener, and Vladlen Koltun. Online continual learning with natural distribution shifts: An empirical study with visual data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8281–8290, 2021.

[Chen *et al.*, 2023] 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.

[Chen *et al.*, 2025] Bojian Chen, Xinmin Zhang, Changqing Shen, Qi Li, and Zhihuan Song. Couda: Continual unsupervised domain adaptation for industrial fault diagnosis under dynamic working conditions. *IEEE Transactions on Industrial Informatics*, 2025.

[Cho *et al.*, 2023] Wonguk Cho, Jinha Park, and Taesup Kim. Complementary domain adaptation and generalization for unsupervised continual domain shift learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11442–11452, 2023.

[De Lange and Tuytelaars, 2021] Matthias De Lange and Tinne Tuytelaars. Continual prototype evolution: Learning online from non-stationary data streams. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8250–8259, 2021.

[Diykh *et al.*, 2016] Mohammed Diykh, Yan Li, and Peng Wen. Eeg sleep stages classification based on time domain features and structural graph similarity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(11):1159–1168, 2016.

[Duan *et al.*, 2023] Tiehang Duan, Zhenyi Wang, Gianfranco Doretto, Fang Li, Cui Tao, and Donald Adjeroh. Replay with stochastic neural transformation for online continual eeg classification. In *2023 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 1874–1879. IEEE, 2023.

[Duan *et al.*, 2024a] Tiehang Duan, Zhenyi Wang, Fang Li, Gianfranco Doretto, Donald A Adjeroh, Yiyi Yin, and Cui Tao. Online continual decoding of streaming eeg signal with a balanced and informative memory buffer. *Neural Networks*, 176:106338, 2024.

[Duan *et al.*, 2024b] Tiehang Duan, Zhenyi Wang, Li Shen, Gianfranco Doretto, Donald A Adjeroh, Fang Li, and Cui Tao. Retain and adapt: Online sequential eeg classification with subject shift. *IEEE Transactions on Artificial Intelligence*, 5(9):4479–4492, 2024.

[Eldele *et al.*, 2021] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

[Herman *et al.*, 2008] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4):317–326, 2008.

[Jenke *et al.*, 2014] Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective computing*, 5(3):327–339, 2014.

[Jia *et al.*, 2020] Ziyu Jia, Youfang Lin, Jing Wang, Ronghao Zhou, Xiaojun Ning, Yuanlai He, and Yaoshuai Zhao. Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *Ijcai*, volume 2021, pages 1324–1330, 2020.

[Jin *et al.*, 2024] Jiarui Jin, Zongnan Chen, Honghua Cai, and Jiahui Pan. Affective eeg-based person identification with continual learning. *IEEE Transactions on Instrumentation and Measurement*, 73:1–16, 2024.

[Khalighi *et al.*, 2016] Sirvan Khalighi, Teresa Sousa, José Moutinho Santos, and Urbano Nunes. Isruc-sleep: A comprehensive public dataset for sleep researchers. *Computer methods and programs in biomedicine*, 124:180–192, 2016.

[Li *et al.*, 2019] Jinpeng Li, Shuang Qiu, Yuan-Yuan Shen, Cheng-Lin Liu, and Huiguang He. Multisource transfer learning for cross-subject eeg emotion recognition. *IEEE transactions on cybernetics*, 50(7):3281–3293, 2019.

[Li *et al.*, 2024a] Dan Li, Hye-Bin Shin, Kang Yin, and Seong-Whan Lee. Domain-incremental learning framework for continual motor imagery eeg classification task. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–5. IEEE, 2024.

[Li *et al.*, 2024b] Qiwei Li, Yuxin Peng, and Jiahuan Zhou. Progressive prototype evolving for dual-forgetting mitigation in non-exemplar online continual learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2477–2486, 2024.

[Li *et al.*, 2025] Dan Li, Hye-Bin Shin, and Kang Yin. Personalized continual eeg decoding: Retaining and transferring knowledge. In *2025 13th International Conference on Brain-Computer Interface (BCI)*, pages 1–4. IEEE, 2025.

[Lopez-Paz and Ranzato, 2017] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

[Mermillod *et al.*, 2013] Martial Mermillod, Aurélia Bugaiska, and Patrick Bonin. The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects, 2013.

[Pan *et al.*, 2023] Tongjie Pan, Yalan Ye, Hecheng Cai, Shudong Huang, Yang Yang, and Guoqing Wang. Multimodal physiological signals fusion for online emotion recognition. In *Proceedings of the 31st ACM international conference on multimedia*, pages 5879–5888, 2023.

[Rolnick *et al.*, 2019] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019.

[Saporta *et al.*, 2022] Antoine Saporta, Arthur Douillard, Tuan-Hung Vu, Patrick Pérez, and Matthieu Cord. Multi-head distillation for continual unsupervised domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3751–3760, 2022.

[Schalk *et al.*, 2004] 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.

[Schulman *et al.*, 2015] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897. PMLR, 2015.

[She *et al.*, 2023a] Qingshan She, Tie Chen, Feng Fang, Jianhai Zhang, Yunyuan Gao, and Yingchun Zhang. Improved domain adaptation network based on wasserstein distance for motor imagery eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1137–1148, 2023.

[She *et al.*, 2023b] Qingshan She, Chenqi Zhang, Feng Fang, Yuliang Ma, and Yingchun Zhang. Multisource associate domain adaptation for cross-subject and cross-session eeg emotion recognition. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023.

[Squire *et al.*, 2015] Larry R Squire, Lisa Genzel, John T Wixted, and Richard G Morris. Memory consolidation. *Cold Spring Harbor perspectives in biology*, 7(8):a021766, 2015.

[Taufique *et al.*, 2021] Abu Md Niamul Taufique, Chowdhury Sadman Jahan, and Andreas Savakis. Conda: Continual unsupervised domain adaptation. *arXiv preprint arXiv:2103.11056*, 2021.

[Taufique *et al.*, 2022] Abu Md Niamul Taufique, Chowdhury Sadman Jahan, and Andreas Savakis. Unsupervised continual learning for gradually varying domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3740–3750, 2022.

[Taufique *et al.*, 2023] Abu Md Niamul Taufique, Chowdhury Sadman Jahan, and Andreas Savakis. Continual unsupervised domain adaptation in data-constrained environments. *IEEE Transactions on Artificial Intelligence*, 5(1):167–178, 2023.

[Wang *et al.*, 2022] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022.

[Wang *et al.*, 2024a] Jing Wang, Xiaojun Ning, Wei Xu, Yunze Li, Ziyu Jia, and Youfang Lin. Multi-source selective graph domain adaptation network for cross-subject eeg emotion recognition. *Neural Networks*, 180:106742, 2024.

[Wang *et al.*, 2024b] Yanshuo Wang, Jie Hong, Ali Cheraghian, Shafin Rahman, David Ahmedt-Aristizabal, Lars Petersson, and Mehrtash Harandi. Continual test-time domain adaptation via dynamic sample selection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1701–1710, 2024.

[Wang *et al.*, 2025] Jing Wang, Zhiyang Feng, Xiaojun Ning, Youfang Lin, Badong Chen, and Ziyu Jia. Two-stream dynamic heterogeneous graph recurrent neural network for multi-label multi-modal emotion recognition. *IEEE Transactions on Affective Computing*, 2025.

[Ye and Bors, 2024] Fei Ye and Adrian G Bors. Online task-free continual generative and discriminative learning via dynamic cluster memory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26202–26212, 2024.

[Zhou *et al.*, 2023] Rushuang Zhou, Zhiguo Zhang, Hong Fu, Li Zhang, Linling Li, Gan Huang, Fali Li, Xin Yang, Yining Dong, Yuan-Ting Zhang, et al. Pr-pl: A novel prototypical representation based pairwise learning framework for emotion recognition using eeg signals. *IEEE Transactions on Affective Computing*, 15(2):657–670, 2023.

[Zhou *et al.*, 2025] Yangxuan Zhou, Sha Zhao, Jiquan Wang, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Brainuicl: An unsupervised individual continual learning framework for eeg applications. In *The Thirteenth International Conference on Learning Representations*, 2025.