SUBJECT-COORDINATE CANONICALIZATION: LEARNING REUSABLE OFFSETS FOR EEG FOUNDATION MODEL ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalography (EEG) foundation models often generalize poorly to unseen subjects, but the structure of subject-dependent variation in their representations remains unclear. Across five EEG foundation models, we compare shared class structure with subject-specific translations and low-rank deformations. Held-out prototype and sample-level reconstruction analyses identify a substantial translation-like component, motivating adaptation through a reusable subject-level offset. Estimating this offset from unlabeled data is challenging because context statistics also reflect class composition. We propose Subject-Coordinate Canonicalization (SCC), which learns to estimate an offset from a small unlabeled context set and subtracts it from disjoint query representations. A permutation-invariant estimator aggregates sample-wise offset proposals, while class-balanced supervision on source subjects reduces sensitivity to context class composition. Leave-one-subject-out response alignment further encourages the corrected representations to preserve task-relevant structure shared across subjects. At inference, the estimated offset is cached and reused, requiring neither target labels nor gradient-based parameter updates. Experiments on three EEG benchmarks spanning emotion recognition, motor imagery, and sleep staging show improvements over source-only models and the evaluated adaptation baselines across all five foundation models. These results demonstrate that learned subject-level translation correction enables effective cross-subject adaptation.

1 INTRODUCTION

Electroencephalography (EEG) foundation models have demonstrated promising transferability across neural decoding tasks through large-scale pretraining Wang et al. (2024); Liu et al. (2026). However, their performance often degrades on unseen subjects, as physiological and recording variations shift the learned representation space Mentzelopoulos et al. (2024). This limits practical deployment, where new users may provide only a few unlabeled calibration samples. Therefore, adapting pretrained EEG representations to unseen subjects without target labels or additional pretraining remains a key challenge.

Existing approaches address cross-subject variability through domain alignment, invariant learning, prototype adaptation, or subject-specific tuning Qiu et al. (2025); Yi et al. (2026); Wang et al. (2025c); Mentzelopoulos et al. (2024). However, they often treat subject variation as an unspecified distribution shift Shen et al. (2025), leaving its underlying geometry unclear. Subjects may introduce complex class-dependent distortions or preserve shared task geometry under a simpler transformation Li et al. (2025). This distinction is critical, since generic invariance may remove task-relevant structure, while overly flexible adaptation can overfit limited target data.

Rather than assuming a particular shift pattern, we diagnose the geometry of subject-conditioned representations extracted from frozen EEG foundation models. As illustrated in Fig. 1, we compare hypotheses with increasing flexibility, ranging from shared class structure alone to subject-wise translation and subject-specific low-rank deformation. Across heterogeneous EEG foundation models Wang et al. (2025b); Zhou et al. (2025); Ma et al. (2026a), subject–class representations preserve shared class organization while exhibiting substantial subject-dependent offsets. Held-out

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.

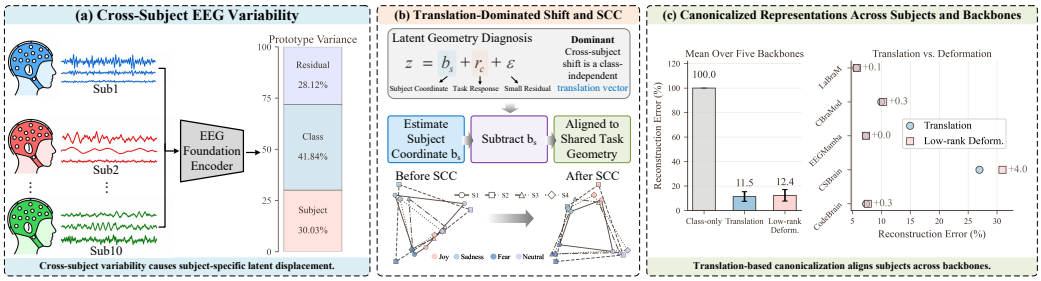

Figure 1: Geometric diagnosis motivating Subject-Coordinate Canonicalization (SCC). (a) EEG foundation representations exhibit subject-dependent latent displacement while preserving shared task structure. (b) Geometric analysis shows that class-independent translation explains most predictable cross-subject variation. (c) Translation-based canonicalization aligns heterogeneous EEG foundation models, whereas flexible deformation yields limited benefits.

reconstruction shows that a single class-independent translation explains most predictable cross-subject variation, whereas additional low-rank deformation provides only marginal improvement. This translation-dominated pattern is further validated at the sample level across diverse decoding paradigms and foundation architectures, indicating that the observation is not an artifact of prototype averaging. Detailed analyses are provided in Appendix A.

This diagnosis suggests a simple adaptation principle: instead of aligning the entire target distribution, estimate a subject-specific coordinate and remove it while preserving shared task responses Chen et al. (2022). For an unseen subject, this coordinate should be inferred from a small unlabeled context set and reused for independent query samples. However, naive context averaging may entangle subject identity with task composition, and query-dependent estimation violates the intended deployment setting. A practical solution should therefore estimate subject coordinates from context alone while maintaining task-relevant geometry.

Motivated by this principle, we propose **Subject-Coordinate Canonicalization** (SCC), a lightweight post-pretraining adaptation framework for EEG foundation representations. SCC uses a permutation-invariant estimator to infer subject coordinates from unlabeled context representations and removes them from disjoint query representations. Training is guided by two complementary objectives: class-balanced coordinate supervision for estimating subject coordinates and leave-one-subject-out response alignment for preserving shared task geometry. Unlike generic centering or distribution alignment, SCC explicitly targets the subject coordinate revealed by geometric diagnosis. At inference, SCC requires neither target labels, gradient updates, nor additional pretraining. Our contributions are fourfold:

- We perform a systematic geometric diagnosis of EEG foundation model representations and reveal that cross-subject variability is dominated by a class-independent translation component while shared task geometry is largely preserved across heterogeneous backbones and decoding paradigms.
- Based on this observation, we formulate cross-subject adaptation as a subject-coordinate recovery problem rather than generic distribution alignment, introducing an additive decomposition of subject-dependent displacement and task-related response.
- We propose Subject-Coordinate Canonicalization (SCC), which estimates a reusable subject coordinate from a small unlabeled context set and canonicalizes query representations through geometry-preserving alignment.
- Extensive experiments across five EEG foundation models and multiple neural decoding paradigms demonstrate consistent improvements on unseen-subject decoding without target labels, test-time optimization, or additional pretraining.

2 Related Work

2.1 EEG Foundation Models

EEG foundation models learn transferable neural representations from large-scale recordings through self-supervised pretraining. Existing models explore diverse strategies, including masked signal modeling, neural tokenization, attention-based architectures, and state-space modeling Jiang et al. (2024); Wang et al. (2025a;b); Zhou et al. (2025); Ma et al. (2026a). Despite substantial progress in representation learning and downstream transfer, most studies focus on pretraining objectives and task performance, while how subject variability is organized within pretrained representations remains largely unexplored. Our work complements these efforts by systematically analyzing subject-conditioned geometry across heterogeneous EEG foundation models and deriving an adaptation strategy from the identified structure.

2.2 Cross-Subject EEG Generalization

Cross-subject EEG decoding has been studied through domain alignment, invariant representation learning, prototype matching, and subject-specific adaptation Ganin et al. (2016); Sun & Saenko (2016); Kobler et al. (2022). Recent test-time adaptation (TTA) approaches further enable adaptation to unseen subjects using only unlabeled target samples. Representative methods optimize model statistics through entropy minimization Wang et al. (2021), update prediction prototypes using target-domain features Iwasawa & Matsuo (2021), or exploit EEG-specific context adaptation strategies Zhang et al. (2025). However, these approaches typically adapt by modifying model parameters or prediction rules without explicitly characterizing the underlying geometry of subject variation. Recent efforts extend adaptation to EEG foundation models through invariant fine-tuning, prototype-conditioned transfer, and source-free adaptation Ma et al. (2026b). Meanwhile, diagnostic studies reveal that pretrained EEG representations retain substantial subject-specific information Lin et al. (2026). Different from previous approaches that directly mitigate distribution discrepancies, we investigate the latent geometry of subject variation and derive adaptation from the discovered transformation structure.

2.3 Representation Geometry and Canonicalization

Representation geometry provides a principled perspective for separating nuisance transformations from task-relevant information. Canonicalization methods estimate transformation-dependent coordinates and map observations into a shared reference space, offering an alternative to generic invariance learning Kaba et al. (2023); Kollár (1997). In EEG decoding, existing geometric approaches mainly focus on covariance alignment, Riemannian representations, or feature-level distribution matching Tibermacine et al. (2026); Chen et al. (2025). However, they typically do not identify how a new subject transforms the shared task representation. Our work first reveals a translation-dominated subject geometry in EEG foundation representations and then learns to estimate and remove this subject coordinate from unlabeled context.

3 Methodology

3.1 Problem Formulation

We consider cross-subject EEG decoding under a subject-disjoint setting. The training set is denoted by $\mathcal{D}_{\mathrm{tr}} = \{(X_i, Y_i, S_i)\}_{i=1}^{N_{\mathrm{tr}}}$, where i indexes the training samples, N_{tr} is the number of training samples, X_i is an EEG segment, $Y_i \in \mathcal{Y}$ is its task label, and $S_i \in \mathcal{S}_{\mathrm{tr}}$ is the corresponding subject identity. The training and test subject sets are disjoint ($\mathcal{S}_{\mathrm{tr}} \cap \mathcal{S}_{\mathrm{te}} = \varnothing$), while sharing the same task-label space $\mathcal{Y}$.

Given an EEG segment $X_i \in \mathbb{R}^{C\times T\times P}$, an EEG foundation encoder f_θ produces a structured channel–temporal representation: $\boldsymbol{H}_i = f_\theta(X_i) \in \mathbb{R}^{C\times T\times d}$, where C, T, and d denote the numbers of channels, temporal patches, and latent dimensions, respectively. A backbone-specific readout p_ω maps the token tensor to the classifier-level representation:

$$\boldsymbol{Z}_i = p_\omega(\boldsymbol{H}_i) \in \mathbb{R}^D, \qquad \boldsymbol{\ell}_i^{\mathrm{base}} = h_\psi(\boldsymbol{Z}_i), \tag{1}$$

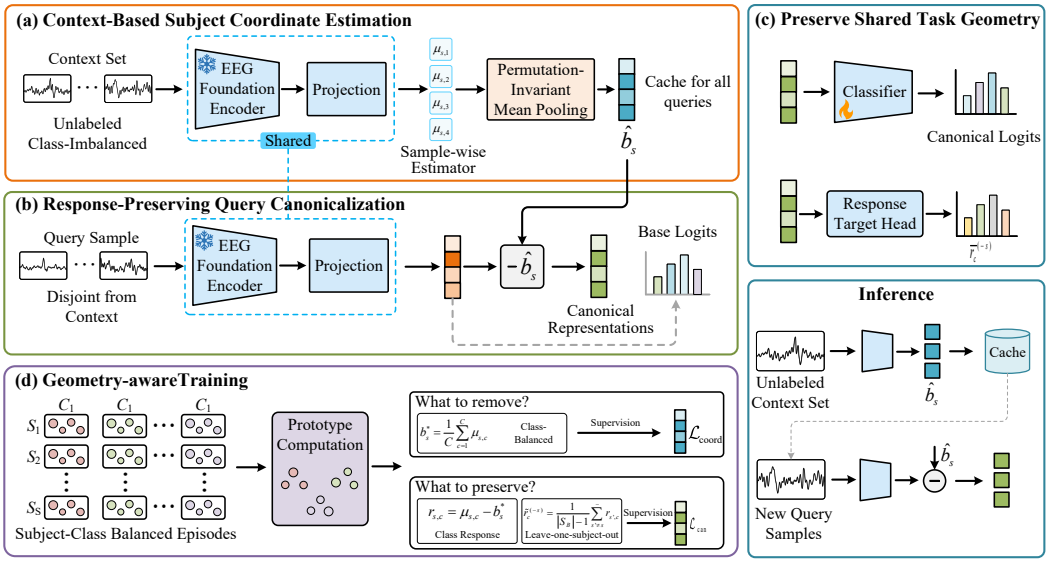

Figure 2: Overview of Subject-Coordinate Canonicalization (SCC). SCC performs geometry-aware adaptation by estimating subject coordinates, removing subject-dependent displacement, and preserving shared task geometry. (a) A permutation-invariant estimator infers a reusable subject coordinate from a small unlabeled context set. (b) The estimated coordinate is subtracted from disjoint query representations and decoded with the shared classifier. (c) Task geometry is preserved through leave-one-subject-out response alignment. (d) Geometry-aware training separates removable subject coordinates from preservable task responses using class-balanced coordinate and response targets. At inference, SCC caches the estimated coordinate without target labels or gradient updates.

where h_ψ is the downstream task classifier. The proposed subject-coordinate canonicalization operates in the representation space of $\boldsymbol{Z}_i$.

For each unseen subject $s \in \mathcal{S}_{\text{te}}$, we assume access to a small unlabeled context set: $\mathcal{C}_s = \left\{X_{s,j}^{\text{ctx}}\right\}_{j=1}^{M_s}$, where M_s denotes the number of context samples. The context set is used to estimate a subject coordinate that is subsequently shared by the query samples of the same subject. Context samples are excluded from query evaluation, and their labels are unavailable to the model at inference. Our objective is to learn a context-conditioned predictor:

$$\widehat{Y}_{s,i}^{\text{qry}} = F_\Theta\left(X_{s,i}^{\text{qry}}; \mathcal{C}_s\right), \tag{2}$$

where Θ denotes the learnable model parameters. The predictor uses neither target labels nor gradient-based adaptation at inference.

3.2 Context-Based Subject Coordinate Estimation

The geometric analysis in Appendix A indicates that the dominant predictable variation across subjects can be characterized as a class-independent translation while the relative task structure is largely preserved. Accordingly, we model a classifier-level representation as:

$$\boldsymbol{Z}_{s,i} = \boldsymbol{b}_s + \boldsymbol{r}_{s,i} + \boldsymbol{\epsilon}_{s,i}, \tag{3}$$

where $\boldsymbol{b}_s \in \mathbb{R}^D$ denotes the subject coordinate, $\boldsymbol{r}_{s,i}$ contains the task-relevant response, and $\boldsymbol{\epsilon}_{s,i}$ represents residual variation.

Given the unlabeled context set $\mathcal{C}_s$, each context sample is first mapped to the classifier-level space: $\boldsymbol{Z}_{s,j}^{\text{ctx}} = p_\omega\left(f_\theta\left(X_{s,j}^{\text{ctx}}\right)\right), j = 1, \ldots, M_s$. A lightweight coordinate estimator g_ϕ then predicts a sample-wise coordinate proposal: $\boldsymbol{u}_{s,j} = g_\phi\left(\boldsymbol{Z}_{s,j}^{\text{ctx}}\right) \in \mathbb{R}^D$.

The proposals from the same subject are aggregated through permutation-invariant mean pooling: $\widehat{\boldsymbol{b}}_s = \frac{1}{M_s}\sum_{j=1}^{M_s} \boldsymbol{u}_{s,j}$. The estimator receives neither context labels nor query representations, yield-

ing a single subject-level coordinate that can be reused across all subsequent queries. We implement g_ϕ as a lightweight nonlinear projection and zero-initialize its final layer. Consequently, $\widehat{\boldsymbol{b}}_s$ is initially zero, so the subsequent canonicalization starts from an identity transformation and gradually learns the subject-dependent correction.

3.3 Response-Preserving Query Canonicalization

Given a query sample $X^{\mathrm{qry}}_{s,i}$, we obtain its classifier-level representation as $\boldsymbol{Z}^{\mathrm{qry}}_{s,i} = p_\omega(f_\theta(X^{\mathrm{qry}}_{s,i}))$. The context-derived coordinate is shared across all queries from subject s and removed through:

$$\widetilde{\boldsymbol{Z}}^{\mathrm{qry}}_{s,i} = \boldsymbol{Z}^{\mathrm{qry}}_{s,i} - \widehat{\boldsymbol{b}}_s, \qquad \boldsymbol{\ell}^{\mathrm{can}}_{s,i} = h_\psi\left(\widetilde{\boldsymbol{Z}}^{\mathrm{qry}}_{s,i}\right). \tag{4}$$

The base and canonical representations share the same classifier h_ψ; hence, SCC changes the representation coordinate rather than introducing a subject-specific decision function. Query samples never participate in estimating the coordinate applied to them.

Subtraction alone does not guarantee a meaningful decomposition, since an unconstrained estimator may absorb task-relevant information into the predicted coordinate. We therefore construct class-balanced episodic targets that encourage the estimated coordinate to capture subject-level variation shared across task classes. Let $\mathcal{I}_{s,c}$ contain all context and query samples from subject s and class c, and define their prototype as $\boldsymbol{\mu}_{s,c} = |\mathcal{I}_{s,c}|^{-1} \sum_{i \in \mathcal{I}_{s,c}} \boldsymbol{Z}_i$. For the shared subject and class sets $\mathcal{S}_\mathcal{B}$ and $\mathcal{Y}_\mathcal{B}$, we compute:

$$\boldsymbol{b}^\star_s = \frac{1}{|\mathcal{Y}_\mathcal{B}|} \sum_{c \in \mathcal{Y}_\mathcal{B}} \boldsymbol{\mu}_{s,c}, \qquad \boldsymbol{r}_{s,c} = \boldsymbol{\mu}_{s,c} - \boldsymbol{b}^\star_s, \qquad \overline{\boldsymbol{r}}^{(-s)}_c = \frac{1}{|\mathcal{S}_\mathcal{B}| - 1} \sum_{\substack{s' \in \mathcal{S}_\mathcal{B} \\ s' \neq s}} \boldsymbol{r}_{s',c}. \tag{5}$$

We interpret $\boldsymbol{b}^\star_s$ as a *class-balanced episodic subject center*. Under the additive model in Eq. equation 3, the prototype-level counterpart satisfies $\boldsymbol{b}^\star_s = \boldsymbol{b}_s + \frac{1}{|\mathcal{Y}_\mathcal{B}|} \sum_{c \in \mathcal{Y}_\mathcal{B}} \boldsymbol{r}_c$. Because the same class subset $\mathcal{Y}_\mathcal{B}$ is shared across subjects within an episode, the second term acts as a common episode-level offset and therefore preserves relative subject coordinates. Equal class weighting removes class-frequency bias, while variation in class coverage is handled through episodic resampling rather than assumed to vanish exactly. The leave-one-subject-out target $\overline{\boldsymbol{r}}^{(-s)}_c$ encourages the canonical representation to retain response structure shared across subjects. All prototype-derived targets are detached during optimization.

To learn subject-coordinate canonicalization, SCC is optimized with two primary objectives: coordinate estimation, which estimates the class-independent subject coordinate from unlabeled context, and response preservation, which encourages canonicalized queries to preserve the shared task-response geometry:

$$\begin{aligned}
\mathcal{L}_{\mathrm{coord}} &= \frac{1}{|\mathcal{S}_\mathcal{B}|} \sum_{s \in \mathcal{S}_\mathcal{B}} \left[\mathrm{SmoothL1}\left(\widehat{\boldsymbol{b}}_s, \boldsymbol{b}^\star_s\right) + \eta_{\mathrm{coord}} \left(1 - \cos\left(\widehat{\boldsymbol{b}}_s, \boldsymbol{b}^\star_s\right)\right) \right], \\
\mathcal{L}_{\mathrm{can}} &= \frac{1}{|\mathcal{Q}_\mathcal{B}|} \sum_{(s,i) \in \mathcal{Q}_\mathcal{B}} \left[1 - \cos\left(\widetilde{\boldsymbol{Z}}^{\mathrm{qry}}_{s,i}, \overline{\boldsymbol{r}}^{(-s)}_{Y_i}\right) + \eta_{\mathrm{can}}\, \mathrm{SmoothL1}\left(\widetilde{\boldsymbol{Z}}^{\mathrm{qry}}_{s,i}, \overline{\boldsymbol{r}}^{(-s)}_{Y_i}\right) \right].
\end{aligned} \tag{6}$$

The coordinate objective supervises the estimated coordinate toward the class-balanced episodic subject center, while the canonicalization objective aligns the corrected query representations with shared task responses from other subjects. The SCC objective is defined as: $\mathcal{L}_{\mathrm{SCC}} = \lambda_{\mathrm{coord}}\mathcal{L}_{\mathrm{coord}} + \lambda_{\mathrm{can}}\mathcal{L}_{\mathrm{can}} + \mathcal{L}^{\mathrm{SCC}}_{\mathrm{reg}}$, where $\mathcal{L}^{\mathrm{SCC}}_{\mathrm{reg}}$ contains lightweight stabilization terms, including proposal consistency, coordinate–response decorrelation, and residual geometric regularization, which are detailed in Appendix B.

3.4 Overall Objective and Inference

All classification losses are evaluated only on the query partition $\mathcal{Q}_\mathcal{B}$. The primary task loss is computed from the canonical logits: $\mathcal{L}_{\mathrm{task}} = \frac{1}{|\mathcal{Q}_\mathcal{B}|} \sum_{(s,i) \in \mathcal{Q}_\mathcal{B}} \mathrm{CE}\left(\boldsymbol{\ell}^{\mathrm{can}}_{s,i}, Y_i\right)$.

We also apply an auxiliary source-space classification loss to the base logits to stabilize the shared readout and classifier: $\mathcal{L}_{\text{base}} = \frac{1}{|\mathcal{Q}_{\mathcal{B}}|} \sum_{(s,i)\in\mathcal{Q}_{\mathcal{B}}} \text{CE}\left(\boldsymbol{\ell}_{s,i}^{\text{base}}, Y_i\right)$. The overall training objective combines query classification and SCC supervision:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_{\text{base}}\mathcal{L}_{\text{base}} + \gamma_t \mathcal{L}_{\text{SCC}}, \tag{7}$$

where $\mathcal{L}_{\text{SCC}}$ combines direct coordinate supervision with objectives that encourage response-preserving canonicalization, and γ_t progressively increases the contribution of these geometric objectives during training. The base loss is used only for training stabilization; inference uses the canonical prediction.

4 Experiments

4.1 Experimental Setup

4.1.1 Datasets

We evaluate SCC on three subject-intensive EEG benchmarks spanning emotion recognition, motor imagery, and sleep staging. All datasets are partitioned by subject, with no subject overlap among the training, validation, and test sets. **FACED** Chen et al. (2023) is a fine-grained affective EEG dataset containing 32-channel recordings from 123 subjects. We use the preprocessed signals standardized to 250 Hz and consider nine emotion categories comprising four positive emotions, four negative emotions, and neutrality. To ensure consistency with the corresponding EEG foundation models, we follow the same subject-wise train/validation/test partitions used in their evaluation protocols. Subjects 1–80, 81–100, and 101–123 are assigned to the training, validation, and test sets, respectively. **PhysioNet-MI** Schalk et al. (2004) contains 64-channel EEG recordings sampled at 160 Hz from 109 subjects. We use the motor-imagery runs and formulate a four-class task involving the left fist, right fist, both fists, and both feet. Subjects 1–70 are used for training, 71–89 for validation, and 90–109 for testing. **ISRUC-Sleep** Khalighi et al. (2016) provides all-night polysomnographic recordings from adults with diverse sleep conditions. We use Subgroup I, which contains one recording from each of 100 subjects, and retain its six EEG channels for five-stage classification over Wake, N1, N2, N3, and REM. Subjects 1–80, 81–90, and 91–100 are used for training, validation, and testing, respectively. These benchmarks differ substantially in acquisition configurations, temporal characteristics, and label spaces, enabling a broad evaluation of unseen-subject adaptation across affective, cognitive, and physiological EEG decoding tasks.

4.1.2 Baselines

We evaluate SCC against source-only foundation models and representative unsupervised adaptation methods. We consider five EEG foundation models: LaBraM Jiang et al. (2024), CBraMod Wang et al. (2025a), EEGMamba Wang et al. (2025b), CSBrain Zhou et al. (2025), and CodeBrain Ma et al. (2026a). For each backbone, we compare the source-only model with EA Wu (2025), ReCenter, Tent Wang et al. (2021), T3A Iwasawa & Matsuo (2021), COME-TTA Zhang et al. (2025) and SCC. All methods share identical data splits, preprocessing, downstream heads, source checkpoints, and target context size. Test-time adaptation baselines are allowed to use only the unlabeled target context and are evaluated on disjoint query samples without target labels. Hyperparameters are selected using validation subjects only.

4.1.3 Metrics and Implementation

We report balanced accuracy and Cohen's kappa as the primary metrics, since EEG decoding tasks often involve class imbalance and multi-class agreement beyond chance. All foundation models are initialized from released pretrained checkpoints and adapted through downstream fine-tuning on training subjects without additional pretraining. Unless otherwise stated, each target subject provides $M = 4$ unlabeled context samples, with all remaining samples used as disjoint queries. The same context-query split is used for all test-time adaptation baselines, and hyperparameters are selected exclusively on validation subjects. Complete optimization details are provided in Appendix D.

Table 1: Overall performance across three EEG decoding tasks and five EEG foundation models. All methods use the same unlabeled target context and are evaluated on disjoint query samples. Results are mean ± standard deviation; bold indicates the best mean within each backbone group.

Methods	FACED, 9-class		PhysioNet-MI, 4-class		ISRUC, 5-class	
	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑
LaBraM	52.73±**1.07**	46.98±1.88	61.73±1.22	49.12±1.92	76.33±1.02	72.31±1.82
+ EA	53.57±1.23	47.44±1.34	63.03±1.15	51.38±1.51	77.85±1.07	74.10±1.54
+ ReCenter	54.66±1.14	49.12±1.25	64.68±1.08	52.60±**1.21**	78.62±0.87	74.91±1.46
+ Tent	52.80±1.25	46.22±1.45	61.87±1.21	49.93±1.67	76.94±0.95	73.18±1.79
+ T3A	54.12±1.16	47.85±1.17	64.43±1.03	52.26±1.46	78.22±0.82	74.83±1.32
+ COME-TTA	54.52±1.21	48.46±1.14	64.75±1.01	52.34±1.37	78.81±0.93	75.17±1.32
+ SCC	**56.39**±1.18	**50.47±1.05**	**66.38±0.92**	**54.36**±1.40	**81.42±0.75**	**78.51±0.97**
CBraMod	55.09±0.89	50.41±1.22	64.17±0.91	52.22±1.69	78.65±1.10	74.42±1.52
+ EA	57.74±0.95	52.27±1.35	64.77±1.03	53.04±1.48	79.07±1.28	74.93±1.49
+ ReCenter	59.08±0.86	53.52±0.95	65.33±0.84	53.12±1.34	79.84±0.94	75.57±1.31
+ Tent	56.43±0.99	51.19±1.36	64.57±1.06	52.90±1.44	78.75±1.15	74.34±1.46
+ T3A	59.59±0.79	54.42±0.82	65.74±0.88	53.67±1.12	80.11±0.84	76.17±1.27
+ COME-TTA	60.52±0.81	55.25±0.85	65.96±0.92	53.95±1.09	80.74±0.87	76.42±1.25
+ SCC	**62.63±0.74**	**57.90±0.75**	**67.21±0.82**	**56.24±1.07**	**82.87±0.81**	**78.32±1.24**
EEGMamba	56.46±0.96	51.34±1.27	63.23±0.84	51.31±1.24	78.99±0.93	74.86±1.22
+ EA	57.25±0.81	51.85±1.18	64.47±0.82	52.29±1.17	79.92±0.89	75.61±1.13
+ ReCenter	58.39±0.86	53.05±0.95	64.98±0.77	52.40±1.09	80.47±0.85	75.66±0.93
+ Tent	56.99±0.84	50.70±1.08	63.42±0.91	51.92±1.26	79.31±0.97	75.10±1.19
+ T3A	58.45±0.79	53.01±0.83	65.19±0.76	53.25±1.10	80.09±0.82	75.85±0.94
+ COME-TTA	58.53±0.82	53.67±0.82	65.69±0.79	53.59±1.08	80.21±0.84	75.73±0.96
+ SCC	**60.84±0.78**	**56.11±0.80**	**67.02±0.74**	**55.35±1.06**	**81.11±0.81**	**77.33±0.89**
CSBrain	57.52±0.42	52.04±**0.36**	63.04±0.90	50.71±1.20	79.25±**0.30**	74.06±1.02
+ EA	58.52±0.51	53.74±0.53	64.38±0.95	52.49±1.14	80.06±0.72	74.24±0.93
+ ReCenter	59.68±0.41	53.69±0.47	65.07±0.87	52.73±0.90	80.15±0.74	75.18±0.72
+ Tent	57.80±0.55	52.21±0.63	63.87±1.02	51.18±1.19	79.38±0.88	74.04±1.08
+ T3A	60.04±0.64	55.19±0.69	65.96±0.82	53.29±0.95	81.07±0.65	75.45±0.74
+ COME-TTA	60.47±0.55	55.71±0.58	65.82±0.84	53.24±0.93	81.53±0.72	75.83±0.77
+ SCC	**62.09±0.37**	**57.49**±0.38	**68.81±0.81**	**56.34±0.87**	**82.88**±0.93	**77.55±0.68**
CodeBrain	59.41±0.98	54.06±0.84	63.86±0.93	50.13±1.32	78.35±**0.33**	74.76±**0.40**
+ EA	60.09±0.91	54.67±0.98	64.55±0.96	51.09±1.28	79.23±1.07	75.78±0.97
+ ReCenter	61.26±0.64	56.14±0.81	65.57±0.90	52.10±1.19	80.58±0.93	75.43±0.99
+ Tent	59.83±0.95	54.57±0.98	64.13±1.04	50.53±1.22	78.81±1.24	75.10±1.27
+ T3A	61.94±0.61	56.60±0.83	65.56±0.91	52.09±1.20	80.88±0.87	75.99±0.89
+ COME-TTA	62.27±0.58	56.96±0.85	65.51±0.95	52.01±1.25	80.82±0.93	75.91±0.91
+ SCC	**64.16±0.50**	**58.97±0.79**	**66.74±0.89**	**53.64±1.18**	**81.97**±0.85	**77.67**±0.58

4.2 Results and Analysis

4.2.1 Unseen-Subject Adaptation Performance

Table 1 reports the results on three representative EEG decoding paradigms, including emotion recognition, motor imagery, and sleep staging. The evaluated EEG foundation models cover diverse architectural paradigms, including Transformer-based and state-space-model-based backbones. SCC consistently improves all five EEG foundation models across all datasets and metrics. Averaged over five backbones, SCC improves the source-only models by $+4.98$, $+4.03$, and $+3.74$ in balanced accuracy on FACED, PhysioNet-MI, and ISRUC, respectively, with corresponding kappa gains of $+5.22$, $+4.49$, and $+3.79$. These results indicate that the proposed subject-coordinate canonicalization is not tied to a specific backbone architecture or task paradigm. Compared with the strongest non-SCC baseline in each setting, SCC further yields average gains of $+1.81$ in balanced accuracy and $+2.12$ in kappa. These consistent improvements demonstrate the effectiveness and generality of SCC for adapting EEG foundation representations to unseen subjects.

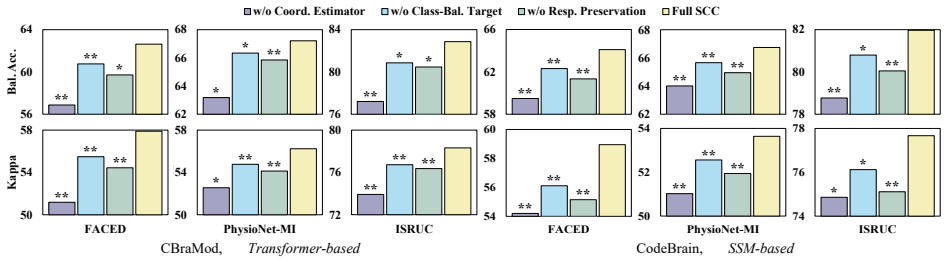

Figure 3: Ablation study of SCC. We remove each core component from Full SCC and report balanced accuracy and Cohen's kappa on CBraMod and CodeBrain across three datasets. Asterisks indicate significant drops from Full SCC: $^{*}p < 0.05$, $^{**}p < 0.01$.

Table 2: Context robustness of SCC using CBraMod as the default backbone. We vary the number of randomly sampled unlabeled context samples and compare diagnostic context compositions at $M = 4$, where M denotes the total number of context samples per unseen subject. Random sampling is used in the default evaluation protocol. The class-diverse oracle and imbalanced settings are diagnostic only: target labels are used by the evaluation script to construct the context composition but are never provided to SCC.

Setting	Context	FACED		PhysioNet-MI		ISRUC	
		Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑
Size	Random, $M = 2$	62.12	56.78	65.72	53.98	80.86	77.20
	Random, $M = 4$	62.63	57.90	67.21	56.24	82.87	78.32
	Random, $M = 8$	62.04	57.10	65.68	54.06	81.12	77.07
	Random, $M = 16$	61.19	56.16	65.33	53.48	80.31	76.67
	Random, $M = 32$	60.25	54.86	64.90	52.93	80.47	76.32
Composition	Class-diverse Oracle, $M = 4$	62.71	58.05	67.54	56.61	82.92	78.35
	Random, $M = 4$	62.63	57.90	67.21	56.24	82.87	78.32
	Imbalanced, $M = 4$	62.26	57.38	65.03	53.29	82.75	77.93

4.2.2 Contribution of SCC Components

Fig. 3 ablates the core components of SCC on two representative foundation backbones, CBraMod and CodeBrain. Full SCC consistently achieves the best balanced accuracy and kappa across both backbones and all datasets. Removing the coordinate estimator leads to the largest degradation, indicating that estimating a reusable subject coordinate is central to unseen-subject adaptation. Removing the class-balanced coordinate target or response-preserving alignment also consistently reduces performance, showing that SCC benefits from both unbiased coordinate supervision and preservation of shared task geometry after canonicalization. These results validate the necessity of the proposed coordinate estimation and response-preserving design.

4.2.3 Robustness to Unlabeled Context

Table 2 evaluates SCC under different context settings using a fixed CBraMod checkpoint. With random unlabeled sampling, SCC achieves strong performance with small contexts, while increasing the context size beyond $M = 4$ does not bring further improvements. This suggests that effective coordinate estimation depends on representative context rather than simply increasing the number of calibration samples. At $M = 4$, the class-diverse oracle is used only as a diagnostic reference constructed with target labels. Random context remains close to this oracle across datasets, indicating that SCC does not rely on target labels or carefully controlled context composition during inference. A fixed-context nested evaluation is further provided in Appendix E.3, showing that the degradation under random scaling mainly results from changes in context composition rather than the increased context size itself.

Table 3: Quantitative effect of canonicalization using CBraMod as the backbone. S-D. and C-D. denote normalized subject-center dispersion and same-class cross-subject dispersion, respectively. G-C. measures class-geometry preservation, and BA denotes balanced accuracy.

Method	FACED				PhysioNet-MI				ISRUC			
	S-D. ↓	C-D. ↓	G-C. ↑	BA ↑	S-D. ↓	C-D. ↓	G-C. ↑	BA ↑	S-D. ↓	C-D. ↓	G-C. ↑	BA ↑
Source-only	1.00	1.00	–	55.09	1.00	1.00	–	64.17	1.00	1.00	–	78.65
ReCenter	0.96	0.98	0.98	59.08	0.99	1.01	0.98	65.33	1.01	1.00	0.98	79.84
SCC	**0.82**	**0.90**	0.99	**62.63**	**0.88**	**0.93**	0.99	**67.21**	**0.92**	**0.96**	0.99	**82.87**

Table 4: Deployment efficiency comparison under the same context-query protocol using CBraMod as the backbone. Grad. indicates whether test-time backpropagation is required. Adapt. and Query denote context adaptation cost per subject and query inference latency, respectively. Runtime excludes one-time setup costs.

Method	Grad.	#Steps	Adapt. ↓	Query ↓	Total ↓	Memory (GB) ↓	Bal.Acc. ↑
EA	×	–	**0.76**	0.85	0.86	1.41	57.74
ReCenter	×	–	1.09	1.07	1.08	**0.39**	59.08
Tent	✓	5	14.44	0.90	1.08	6.16	56.43
T3A	×	–	36.10	0.95	1.40	1.41	59.59
SCC	×	–	1.03	**0.76**	**0.77**	0.41	**62.63**

4.2.4 Verifying Subject-Coordinate Canonicalization

To verify whether SCC operates according to the proposed geometric principle, we examine two expected effects of canonicalization: reducing subject-dependent displacement while preserving task-relevant class geometry. As shown in Table 3, SCC consistently decreases subject-center dispersion (S-D.) and same-class cross-subject dispersion (C-D.) across all three datasets, while maintaining a high geometry correlation (G-C.) of 0.99. Although ReCenter also performs context-based subtraction during inference, it only provides limited reduction in subject dispersion and may even increase cross-subject dispersion in certain cases. This indicates that SCC is not merely a feature recentering operation; instead, it selectively removes subject-related displacement while preserving shared task geometry through coordinate estimation and response-preserving constraints. The consistent improvement in balanced accuracy further demonstrates that the desired geometric effect translates into better unseen-subject decoding.

4.2.5 Optimization-Free Adaptation Efficiency

Table 4 compares the deployment efficiency of SCC with representative adaptation methods under the same context-query protocol. SCC achieves the best decoding performance without test-time gradient optimization. It estimates the subject coordinate once from unlabeled context and reuses it for subsequent queries, introducing only 1.03 ms per subject for adaptation and 0.76 ms per query for inference. In contrast, Tent requires iterative gradient updates and incurs substantially higher memory consumption, while T3A introduces additional context adaptation overhead due to prototype refinement. These results demonstrate that SCC enables effective unseen-subject adaptation with low deployment cost, avoiding the optimization burden of conventional test-time adaptation.

5 Conclusion

We reveal that cross-subject variability in EEG foundation representations is dominated by class-independent displacement while preserving shared task geometry. Based on this insight, we propose Subject-Coordinate Canonicalization (SCC), which estimates and removes subject coordinates from unlabeled context representations. Extensive experiments across diverse EEG datasets and foundation models demonstrate improved unseen-subject decoding without target labels, additional pretraining, or gradient-based adaptation. SCC provides a simple geometry-driven perspective for separating subject-specific variation from task-relevant neural structure.

REFERENCES

Huayu Chen, Junxiang Li, Huanhuan He, Jing Zhu, Shuting Sun, Xiaowei Li, and Bin Hu. Toward the construction of affective brain-computer interface: A systematic review. *ACM Computing Surveys*, 57(6):1–56, 2025.

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.

Xun Chen, Chang Li, Aiping Liu, Martin J. McKeown, Ruobing Qian, and Z. Jane Wang. Toward open-world electroencephalogram decoding via deep learning: A comprehensive survey. *IEEE Signal Processing Magazine*, 39:117–134, 2022.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17:1–35, 2016.

Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2427–2440, 2021.

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.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pp. 15546–15566. PMLR, 2023.

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.

Reinmar Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. Spd domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg. In *Advances in Neural Information Processing Systems*, volume 35, pp. 6219–6235. Curran Associates, Inc., 2022.

János Kollár. Quotient spaces modulo algebraic groups. *Annals of mathematics*, 145(1):33–79, 1997.

Dongdong Li, Shengyao Huang, Li Xie, Zhe Wang, and Jiazhen Xu. Neuron perception inspired eeg emotion recognition with parallel contrastive learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36:14049–14062, 2025.

Jun-You Lin, Ying Choon Wu, and Tzyy-Ping Jung. The identity trap in eeg foundation models: A diagnostic audit. *arXiv preprint arXiv:2606.06647*, 2026.

Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, and Dongrui Wu. Eeg foundation models: Progresses, benchmarking, and open problems. *arXiv preprint arXiv:2601.17883*, 2026.

Jingying Ma, Feng Wu, Qika Lin, Yucheng Xing, Chenyu Liu, Ziyu Jia, and Mengling Feng. Codebrain: Bridging decoupled tokenizer and multi-scale architecture for EEG foundation model. In *The Fourteenth International Conference on Learning Representations*, 2026a.

Jingying Ma, Feng Wu, Yucheng Xing, Qika Lin, Tianyu Liu, Chenyu Liu, Ziyu Jia, and Mengling Feng. Scope: Structured prototype-guided adaptation for eeg foundation models with limited labels. *arXiv preprint arXiv:2602.17251*, 2026b.

Georgios Mentzelopoulos, Evangelos Chatzipantazis, Ashwin G. Ramayya, Michelle J. Hedlund, Vivek P. Buch, Kostas Daniilidis, Konrad P. Kording, and Flavia Vitale. Neural decoding from stereotactic eeg: accounting for electrode variability across subjects. In *Advances in Neural Information Processing Systems*, volume 37, pp. 108600–108624. Curran Associates, Inc., 2024.

Lina Qiu, Zuorui Ying, Xianyue Song, Weisen Feng, Chengju Zhou, and Jiahui Pan. Mtada: A multi-task adversarial domain adaptation network for eeg-based cross-subject emotion recognition. *IEEE Transactions on Affective Computing*, 16:3354–3368, 2025.

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.

Jian Shen, Lechun You, Yu Ma, Zeguang Zhao, Huajian Liang, Yanan Zhang, and Bin Hu. Uadaan: An uncertainty-aware dynamic adversarial adaptation network for eeg-based depression recognition. *IEEE Transactions on Affective Computing*, 16:2130–2141, 2025.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pp. 443–450. Springer, 2016.

Imad Eddine Tibermacine, Samuele Russo, and Christian Napoli. Orthogonal basis transformation layers for eeg-bci: Geometry-preserving mid-level information fusion on spd manifolds. *Information Fusion*, pp. 104445, 2026.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.

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 *Advances in Neural Information Processing Systems*, volume 37, pp. 39249–39280. Curran Associates, Inc., 2024.

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. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 75310–75346, 2025a.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Shijian Li, and Gang Pan. Eegmamba: An eeg foundation model with mamba. *Neural Networks*, pp. 107816, 2025b.

Qiang Wang, Liying Yang, Qian Zhang, Jingtao Du, and Yumeng Ye. Fine-grained label propagation via density-based prototype matching for cross-subject eeg emotion recognition. *Knowledge-Based Systems*, pp. 114650, 2025c.

Dongrui Wu. Revisiting euclidean alignment for transfer learning in eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 22(3):031005, 2025.

Yufan Yi, Yan Tian, Yiping Xu, and Bo Yang. Hit: Hyperbolic transformer with identity-invariant representation for cross-subject eeg emotion recognition. *IEEE Transactions on Affective Computing*, pp. 1–16, 2026.

Qingyang Zhang, Yatao Bian, Xinke Kong, Peilin Zhao, and Changqing Zhang. Come: Test-time adaption by conservatively minimizing entropy. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 18506–18535, 2025.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. Csbrain: A cross-scale spatiotemporal brain foundation model for eeg decoding. In *Advances in Neural Information Processing Systems*, volume 38, pp. 87150–87195. Curran Associates, Inc., 2025.

A Details of Geometric Analysis

This section provides additional details and dataset-wise results for the geometric analysis used to motivate the additive subject-coordinate model in Section 3.2. All analyses are conducted on classifier-level representations extracted before canonicalization.

A.1 PROTOTYPE CONSTRUCTION AND GEOMETRIC HYPOTHESES

For subject s and class c, let $\mathcal{I}_{s,c} = \{i \mid S_i = s, Y_i = c\}$ and $N_{s,c} = |\mathcal{I}_{s,c}|$. We construct the subject–class prototype as:

$$\boldsymbol{\mu}_{s,c} = \frac{1}{N_{s,c}} \sum_{i \in \mathcal{I}_{s,c}} \boldsymbol{Z}_i. \tag{8}$$

To avoid confounding subject effects with missing class coverage, we retain only subjects with a complete set of class prototypes. We then compare three geometric hypotheses of increasing flexibility:

$$\begin{aligned} \mathcal{M}_{\text{cls}}: &\quad \boldsymbol{\mu}_{s,c} = \boldsymbol{r}_c + \boldsymbol{\epsilon}_{s,c}, \\ \mathcal{M}_{\text{trans}}: &\quad \boldsymbol{\mu}_{s,c} = \boldsymbol{b}_s + \boldsymbol{r}_c + \boldsymbol{\epsilon}_{s,c}, \\ \mathcal{M}_{\text{lr}}: &\quad \boldsymbol{\mu}_{s,c} = \boldsymbol{b}_s + T_{\boldsymbol{d}_s}(\boldsymbol{r}_c) + \boldsymbol{\epsilon}_{s,c}, \end{aligned} \tag{9}$$

where $\boldsymbol{r}_c \in \mathbb{R}^D$ denotes the shared class response and $\boldsymbol{b}_s \in \mathbb{R}^D$ denotes a class-independent subject coordinate. This prototype-level formulation is the class-averaged counterpart of the sample-level decomposition in Eq. 3, where the sample-specific response is aggregated into a shared class response: $\boldsymbol{r}_{s,i} \to \boldsymbol{r}_c$. For the low-rank hypothesis, we define:

$$T_{\boldsymbol{d}_s}(\boldsymbol{r}) = (\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\boldsymbol{r} + \mathbf{U}\,\mathrm{diag}(\boldsymbol{d}_s)\mathbf{U}^\top \boldsymbol{r}, \tag{10}$$

where $\mathbf{U} \in \mathbb{R}^{D \times R}$ is a shared orthonormal basis ($\mathbf{U}^\top\mathbf{U} = \mathbf{I}$, $R \ll D$) and $\boldsymbol{d}_s \in \mathbb{R}^R_+$ contains subject-specific scaling coefficients.

Thus, $\mathcal{M}_{\text{lr}}$ allows the shared class geometry to undergo subject-dependent deformation only within an R-dimensional subspace, while leaving its orthogonal component unchanged. Importantly, when $\boldsymbol{d}_s = \mathbf{1}$, $T_{\boldsymbol{d}_s}(\boldsymbol{r}) = \boldsymbol{r}$, so $\mathcal{M}_{\text{lr}}$ reduces exactly to $\mathcal{M}_{\text{trans}}$. We impose $\sum_c \boldsymbol{r}_c = \mathbf{0}$ to resolve the translation ambiguity between $\boldsymbol{b}_s$ and $\boldsymbol{r}_c$.

A.2 IMPLEMENTATION DETAILS OF LOW-RANK DEFORMATION ANALYSIS

For each held-out evaluation split, we fit the shared canonical class geometry and low-rank deformation basis using only the training subjects. We evaluate deformation ranks $R \in \{1, 2, 4, 8, 16, 32\}$, excluding ranks larger than the representation dimension. The low-rank model is optimized with AdamW for 1,000 steps using a learning rate of 2×10^{-2} and weight decay of 10^{-4}. Subject-specific scaling coefficients are constrained to $[0.25, 4.0]$, with an additional log-scale regularization weight of 10^{-3}.

For each held-out subject, only the context-class prototypes are used to estimate its subject offset and low-rank scaling coefficients. The latter are obtained by five alternating closed-form updates with ridge regularization 10^{-3}. The estimated transformation is then applied to query classes that are strictly excluded from subject-specific parameter estimation. We hold out 20% of subjects in each split and repeat the subject/class partitioning 20 times. All shared quantities, including the canonical class responses and low-rank basis, are re-estimated independently for every split. For visualization and quantitative comparison, we report the low-rank deformation result with the rank achieving the lowest reconstruction error among the evaluated candidates. This provides a best-case estimate of the explanatory capacity of low-rank subject–class interactions, rather than a tuned adaptation model.

A.3 HELD-OUT EVALUATION

We evaluate each hypothesis by leaving out one subject–class prototype and reconstructing it from the remaining grid. Let $\widehat{\boldsymbol{\mu}}_{s,c}$ denote the reconstructed prototype and $\mathcal{G}$ the evaluated prototype grid. We report the normalized reconstruction error:

$$E_{\text{rec}} = \frac{\sum_{(s,c)\in\mathcal{G}} \|\boldsymbol{\mu}_{s,c} - \widehat{\boldsymbol{\mu}}_{s,c}\|_2^2}{\sum_{(s,c)\in\mathcal{G}} \|\boldsymbol{\mu}_{s,c} - \overline{\boldsymbol{\mu}}\|_2^2}, \tag{11}$$

where $\overline{\boldsymbol{\mu}}$ is the global prototype mean. We additionally measure response angular error:

$$E_{\text{ang}} = \frac{1}{|\mathcal{G}|} \sum_{(s,c)\in\mathcal{G}} \left[1 - \cos\left(\boldsymbol{\mu}_{s,c} - \boldsymbol{b}_s, \overline{\boldsymbol{r}}_c^{(-s)}\right)\right], \tag{12}$$

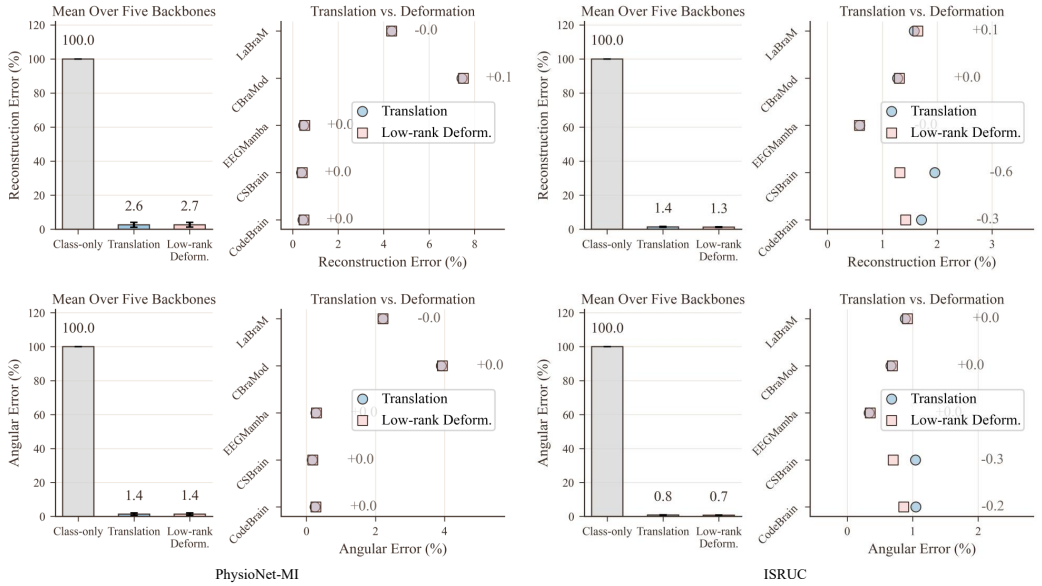

Figure 4: Additional geometric analysis on PhysioNet-MI and ISRUC. Top: normalized reconstruction error; bottom: response angular error. The class-only hypothesis is normalized to 100% for each dataset and metric. Across both datasets, adding a subject-specific translation sharply reduces both errors, while low-rank deformation provides only marginal additional improvement, supporting the translation-dominated subject-coordinate assumption underlying SCC.

where $\overline{\boldsymbol{r}}_c^{(-s)}$ denotes the leave-one-subject-out class response. Lower reconstruction and angular errors indicate better explanatory power and response-geometry preservation. For visualization, reconstruction errors are reported relative to the class-only hypothesis, which is normalized to 100%.

A.4 Additional Evidence for Translation-Dominated Subject Shift

We further examine whether the translation-dominated geometry observed in the main text is consistent across additional EEG decoding paradigms. Fig. 4 reports two complementary held-out prototype metrics on PhysioNet-MI and ISRUC: normalized reconstruction error, which measures how well each geometric hypothesis explains subject–class prototypes, and response angular error, which measures whether the class-response geometry remains aligned after removing the estimated subject component. Across both datasets and both metrics, the class-only model yields substantially larger error, while introducing a subject-specific translation reduces the error to a small fraction of the class-only baseline. On PhysioNet-MI, translation and low-rank deformation achieve nearly identical average reconstruction and response errors over five backbones. On ISRUC, the more flexible low-rank deformation model provides only a marginal additional reduction, decreasing the average reconstruction error from 1.4% to 1.3% and the response angular error from 0.8% to 0.7%. These results indicate that a class-independent translation captures the dominant predictable component of cross-subject variation while preserving response geometry, supporting the additive subject-coordinate model used by SCC.

A.5 Sample-Level Geometry Analysis

Table 5 further examines whether the translation-dominated structure is merely induced by subject–class prototype averaging. Unlike the prototype-level analysis, this experiment evaluates held-out reconstruction on individual samples, which retain trial-level variability, within-class heterogeneity, and nonstationary noise. Across all datasets and backbones, the subject-translation model substantially reduces both reconstruction error and response angular error relative to the class-only baseline, which is normalized to 100 in each setting. For example, the average reconstruction error of translation is reduced to 46.05, 18.64, and 13.52 on FACED, PhysioNet-MI, and ISRUC, respectively,

Table 5: Sample-level geometry analysis across datasets and backbones. We report held-out reconstruction error and response angular error on individual samples. All errors are normalized by the class-only hypothesis within each dataset–backbone setting, so the omitted class-only baseline equals 100. T denotes the subject-translation model, LR denotes the low-rank deformation model, and Δ denotes LR minus T; negative values indicate lower error for LR.

Dataset	Backbone	Reconstruction Error ↓			Angular Error ↓		
		T	LR	Δ	T	LR	Δ
FACED	LaBraM	31.03	31.04	+0.01	17.52	17.53	+0.01
	CBraMod	49.20	49.38	+0.18	19.25	20.01	+0.76
	EEGMamba	38.17	38.18	+0.01	20.28	20.30	+0.02
	CSBrain	76.09	76.94	+0.85	21.95	29.77	+7.82
	CodeBrain	35.74	35.99	+0.25	20.54	20.71	+0.17
	Avg.	46.05	46.31	+0.26	19.91	21.66	+1.75
PhysioNet-MI	LaBraM	40.78	40.81	+0.03	23.04	23.06	+0.02
	CBraMod	38.95	38.98	+0.03	22.18	22.20	+0.02
	EEGMamba	5.40	5.44	+0.04	3.23	3.24	+0.01
	CSBrain	4.11	4.14	+0.03	1.85	1.87	+0.02
	CodeBrain	3.97	4.05	+0.08	2.22	2.26	+0.04
	Avg.	18.64	18.68	+0.04	10.50	10.53	+0.03
ISRUC	LaBraM	31.94	31.88	-0.06	19.92	19.89	-0.03
	CBraMod	18.71	18.72	+0.01	9.83	9.84	+0.01
	EEGMamba	3.72	3.67	-0.05	2.46	2.45	-0.01
	CSBrain	6.79	6.11	-0.68	3.76	3.37	-0.39
	CodeBrain	6.46	6.15	-0.31	4.04	3.86	-0.18
	Avg.	13.52	13.31	-0.21	8.00	7.88	-0.12

with corresponding angular errors of 19.91, 10.50, and 8.00. In contrast, the more flexible low-rank deformation model provides no consistent additional benefit: it slightly increases the average error on FACED and PhysioNet-MI, and only marginally reduces it on ISRUC. These results indicate that subject-wise translation captures a dominant predictable component of cross-subject variation even at the sample level, while the remaining residuals are better attributed to trial-specific variability rather than systematic subject-specific deformation.

B Additional Details of Subject-Coordinate Canonicalization

B.1 Estimator Parameterization

The coordinate estimator is implemented as $g_\phi(\boldsymbol{Z}) = \boldsymbol{W}_2\sigma\left(\boldsymbol{W}_1\,\mathrm{LN}\left(\boldsymbol{Z}\right) + \boldsymbol{a}_1\right) + \boldsymbol{a}_2$. The final projection parameters $\boldsymbol{W}_2$ and $\boldsymbol{a}_2$ are initialized to zero, so the predicted coordinate is initially zero and canonicalization begins as an identity mapping. The estimator is applied independently to each context representation before subject-wise mean aggregation and receives neither task labels nor query representations.

B.2 Coordinate Regularization

In addition to the direct coordinate and canonical-response objectives in Section 3.3, we introduce lightweight regularizers to stabilize context-based coordinate estimation and discourage the estimated coordinate from retaining task-related variation.

First, proposal consistency encourages sample-wise proposals from the same subject to agree with their aggregated coordinate:

$$\mathcal{L}_{\mathrm{cons}} = \frac{1}{|\mathcal{S}_{\mathcal{B}}|}\sum_{s\in\mathcal{S}_{\mathcal{B}}}\frac{1}{M_s}\sum_{j=1}^{M_s}\left\|\boldsymbol{u}_{s,j} - \widehat{\boldsymbol{b}}_s\right\|_2^2. \tag{13}$$

We further reduce linear dependence between the removed coordinate and the remaining canonical representation through:

$$\mathcal{L}_{\mathrm{cov}} = \frac{1}{D^2} \left\| \frac{1}{N_q - 1} \boldsymbol{B}_{\mathrm{c}}^{\top} \widetilde{\boldsymbol{Z}}_{\mathrm{c}} \right\|_{\mathrm{F}}^{2}, \tag{14}$$

where $\boldsymbol{B}_{\mathrm{c}} \in \mathbb{R}^{N_q \times D}$ and $\widetilde{\boldsymbol{Z}}_{\mathrm{c}} \in \mathbb{R}^{N_q \times D}$ denote the query-associated subject coordinates and canonical query representations after batch centering, respectively. Together, $\mathcal{L}_{\mathrm{cons}}$ and $\mathcal{L}_{\mathrm{cov}}$ encourage the estimated coordinate to be stable within a subject and less entangled with the retained canonical representation.

B.3 Residual Geometry Regularization

We additionally regularize the geometry that remains after canonicalization. Let $\mathcal{E}_{\mathcal{B}} = \mathcal{C}_{\mathcal{B}} \cup \mathcal{Q}_{\mathcal{B}}$ denote all samples in an episode, and define:

$$\widetilde{\boldsymbol{Z}}_{s,i} = \boldsymbol{Z}_{s,i} - \widehat{\boldsymbol{b}}_s. \tag{15}$$

For each subject, we compute its canonical center:

$$\widetilde{\boldsymbol{\mu}}_s = \frac{1}{|\mathcal{E}_s|} \sum_{i \in \mathcal{E}_s} \widetilde{\boldsymbol{Z}}_{s,i}, \qquad \widetilde{\boldsymbol{\mu}} = \frac{1}{|\mathcal{S}_{\mathcal{B}}|} \sum_{s \in \mathcal{S}_{\mathcal{B}}} \widetilde{\boldsymbol{\mu}}_s. \tag{16}$$

Residual subject-center dispersion is penalized by:

$$\mathcal{L}_{\mathrm{ctr}} = \frac{1}{|\mathcal{S}_{\mathcal{B}}|} \sum_{s \in \mathcal{S}_{\mathcal{B}}} \|\widetilde{\boldsymbol{\mu}}_s - \widetilde{\boldsymbol{\mu}}\|_2^2. \tag{17}$$

We also construct canonical subject–class prototypes $\widetilde{\boldsymbol{\mu}}_{s,c}$ and define their centered responses as:

$$\widetilde{\boldsymbol{r}}_{s,c} = \widetilde{\boldsymbol{\mu}}_{s,c} - \frac{1}{|\mathcal{Y}_{\mathcal{B}}|} \sum_{c' \in \mathcal{Y}_{\mathcal{B}}} \widetilde{\boldsymbol{\mu}}_{s,c'}. \tag{18}$$

Their cross-subject consistency is encouraged through:

$$\mathcal{L}_{\mathrm{rsp}} = \frac{1}{|\mathcal{S}_{\mathcal{B}}||\mathcal{Y}_{\mathcal{B}}|} \sum_{s \in \mathcal{S}_{\mathcal{B}}} \sum_{c \in \mathcal{Y}_{\mathcal{B}}} d_{\mathrm{rsp}} \left(\widetilde{\boldsymbol{r}}_{s,c}, \overline{\boldsymbol{r}}_{c}^{(-s)} \right), \tag{19}$$

where $d_{\mathrm{rsp}}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \cos(\boldsymbol{x}, \boldsymbol{y}) + \eta_{\mathrm{rsp}} \operatorname{SmoothL1}(\boldsymbol{x}, \boldsymbol{y})$. The leave-one-subject-out response $\overline{\boldsymbol{r}}_{c}^{(-s)}$ is treated as a detached geometric target, consistent with Section 3.3.

Importantly, $\mathcal{L}_{\mathrm{rsp}}$ does not directly supervise the coordinate estimator. Because the same coordinate $\widehat{\boldsymbol{b}}_s$ is subtracted from every class prototype of subject s, it cancels under within-subject centering:

$$\begin{aligned} \widetilde{\boldsymbol{r}}_{s,c} &= \left(\boldsymbol{\mu}_{s,c} - \widehat{\boldsymbol{b}}_s \right) - \frac{1}{|\mathcal{Y}_{\mathcal{B}}|} \sum_{c'} \left(\boldsymbol{\mu}_{s,c'} - \widehat{\boldsymbol{b}}_s \right) \\ &= \boldsymbol{\mu}_{s,c} - \boldsymbol{b}_s^{\star}. \end{aligned} \tag{20}$$

Therefore, $\mathcal{L}_{\mathrm{rsp}}$ acts as a representation-level geometry regularizer, encouraging the trainable readout to preserve cross-subject consistency of centered task responses, rather than directly updating g_ϕ. In contrast, $\mathcal{L}_{\mathrm{ctr}}$ retains explicit dependence on $\widehat{\boldsymbol{b}}_s$ and can directly regularize the estimated subject coordinate.

The complete SCC regularization term is:

$$\mathcal{L}_{\mathrm{reg}}^{\mathrm{SCC}} = \lambda_{\mathrm{cons}} \mathcal{L}_{\mathrm{cons}} + \lambda_{\mathrm{cov}} \mathcal{L}_{\mathrm{cov}} + \lambda_{\mathrm{ctr}} \mathcal{L}_{\mathrm{ctr}} + \lambda_{\mathrm{rsp}} \mathcal{L}_{\mathrm{rsp}}. \tag{21}$$

C Optimization and Episodic Protocol

C.1 Progressive Schedule

To avoid imposing strong geometric constraints before the task-specific readout becomes sufficiently discriminative, we progressively activate the SCC objective using:

$$\rho(t; T_{\mathrm{w}}, T_{\mathrm{r}}) = \mathrm{Clip}\left(\frac{t - T_{\mathrm{w}}}{T_{\mathrm{r}}}, 0, 1 \right), \qquad T_{\mathrm{r}} > 0. \tag{22}$$

Table 6: Implementation hyperparameters across datasets and backbones. C_{tok} and T_{tok} denote the number of channel and temporal tokens used by the backbone wrapper. G/B and S/G denote the number of condition groups per batch and subjects per group, respectively.

Dataset	Backbone	Epochs	LR	WD	Clip	$C_{\mathrm{tok}}/T_{\mathrm{tok}}$	G/B, S/G
FACED	LaBraM	60	1×10^{-4}	5×10^{-2}	1	32 / 10	4, 4
	CBraMod	60	1×10^{-4}	5×10^{-2}	1	32 / 10	4, 4
	EEGMamba	60	1×10^{-4}	5×10^{-2}	1	32 / 10	4, 4
	CSBrain	60	1×10^{-4}	5×10^{-2}	1	32 / 10	4, 4
	CodeBrain	60	5×10^{-5}	5×10^{-4}	5	32 / 10	4, 4
PhysioNet-MI	LaBraM	40	1×10^{-4}	5×10^{-2}	1	64 / 4	4, 8
	CBraMod	40	1×10^{-4}	5×10^{-2}	1	64 / 4	4, 8
	EEGMamba	40	1×10^{-4}	5×10^{-2}	1	64 / 4	4, 8
	CSBrain	40	1×10^{-4}	5×10^{-2}	1	64 / 4	4, 8
	CodeBrain	40	5×10^{-5}	5×10^{-4}	5	64 / 4	4, 8
ISRUC	LaBraM	20	1×10^{-4}	5×10^{-2}	1	6 / 30	5, 6
	CBraMod	20	1×10^{-4}	5×10^{-2}	1	6 / 30	5, 6
	EEGMamba	20	1×10^{-4}	5×10^{-2}	1	6 / 30	5, 6
	CSBrain	20	1×10^{-4}	5×10^{-2}	1	6 / 30	5, 6
	CodeBrain	20	5×10^{-5}	5×10^{-4}	5	6 / 30	5, 6

We set $\gamma_t = \gamma_{\max}\rho(t; T_\gamma^{\mathrm{w}}, T_\gamma^{\mathrm{r}})$, which gradually increases the contribution of $\mathcal{L}_{\mathrm{SCC}}$ during training. Together with the zero initialization of the final layer of g_ϕ, this makes SCC start from an identity canonicalization ($\widehat{\boldsymbol{b}}_s \approx \mathbf{0}$) and progressively introduce subject-dependent correction.

C.2 Classification Supervision

The training objective in Eq. equation 7 is applied directly. The primary classification loss is evaluated on canonical query logits, while the auxiliary base classification loss is evaluated on the corresponding uncanonicalized logits. Both losses use only query samples. Context samples are used for coordinate estimation and for constructing detached geometric targets during training, but never contribute to query evaluation.

C.3 Episodic Training and Inference

Each training episode samples a shared subset of task classes and a set of subjects for each selected class. The same class subset $\mathcal{Y}_{\mathcal{B}}$ is shared by all participating subjects within an episode. For each selected subject–class pair, two samples are drawn, with one assigned to the context partition and the other to the disjoint query partition. Only context representations are provided to the coordinate estimator. For FACED, four classes are sampled from the nine-class label space in each episode and the subset is resampled across episodes. PhysioNet-MI and ISRUC use all four and five task classes, respectively. Training labels are used only by the episodic sampler and to construct prototype-based geometric supervision; they are never provided as inputs to the coordinate estimator.

For an unseen subject, SCC encodes the unlabeled context set once and caches:

$$\widehat{\boldsymbol{b}}_s = \frac{1}{M_s}\sum_{j=1}^{M_s} g_\phi\left(\boldsymbol{Z}_{s,j}^{\mathrm{ctx}}\right). \tag{23}$$

The cached coordinate is reused for all subsequent query samples from that subject. Inference requires neither target labels nor gradient-based parameter updates, and query samples never participate in estimating the coordinate applied to them.

D Implementation Details

D.1 Training Details

All experiments are implemented in PyTorch. For each dataset, we follow the same subject-disjoint train/validation/test split for all methods. Unless otherwise specified, models are optimized with AdamW, dropout rate 0.1, and gradient clipping. The default context size for validation and testing is $M = 4$ unlabeled samples per unseen subject. The same context-query split is used for all context-aware methods, and context samples are excluded from query evaluation. Detailed implementation hyperparameters are summarized in Table 6.

During SCC training, the condition grouping variable is set to the task label, such that each condition group corresponds directly to a task class. Each episode samples a shared subset of task classes, and the same class subset $\mathcal{Y}_{\mathcal{B}}$ is used for all participating subjects within that episode. For every selected subject–class pair, two samples are drawn: one is assigned to the context partition and the other to the disjoint query partition. Training labels are used only by the episodic sampler and prototype-based supervision; they are never provided as inputs to the coordinate estimator. The episode composition is dataset dependent. For FACED, four classes are sampled from the nine-class label space in each episode, with four subjects per selected class. For PhysioNet-MI, all four task classes are included with eight subjects per class, while ISRUC includes all five classes with six subjects per class. For FACED, the four-class subset is resampled across episodes so that different portions of the label space contribute throughout training. Consequently, class-coverage variation occurs primarily in FACED, whereas PhysioNet-MI and ISRUC use the complete task label set in each episode.

The SCC objective combines the query classification loss, an auxiliary base classification loss, coordinate supervision, canonical response alignment, offset consistency, and coordinate–response decorrelation. We set the coordinate estimator hidden dimension to 256, use $\lambda_{\mathrm{coord}} = 0.1$, $\lambda_{\mathrm{can}} = 0.1$, $\lambda_{\mathrm{cons}} = 0.1$, and $\lambda_{\mathrm{cov}} = 0.1$ for all backbones. All inference-time SCC predictions are made without target labels, test-time gradient updates, or query samples for coordinate estimation.

D.2 Baseline Implementation Details

All baselines are evaluated under the same subject-disjoint protocol and use the same train, validation, and test subjects as SCC. For fairness, context-aware baselines use the same unlabeled context set as SCC, and all methods are evaluated on the same disjoint query samples. No query sample is used to estimate context statistics or adapt model parameters.

ReCenter. ReCenter estimates the target-subject shift by directly averaging the base representations of unlabeled context samples and subtracts this empirical mean from query representations. It requires neither target labels nor test-time gradient updates. This baseline evaluates whether simple context centering is sufficient for subject canonicalization. However, under the additive decomposition, the context mean may contain not only the subject coordinate but also the average task response induced by the sampled context composition. In contrast, SCC learns a coordinate estimator to extract subject-dependent displacement from context representations and uses geometric supervision to encourage coordinate–response separation while preserving task-related structure.

EA. Euclidean Alignment estimates the target-subject covariance statistics from unlabeled context samples and applies the resulting alignment transform to query samples from the same subject. The transform is estimated once per subject and does not use query labels.

Tent. Tent adapts the model on the unlabeled context samples by minimizing prediction entropy. Following the context-query protocol, adaptation is performed only on the context set and is reset for each target subject before query evaluation.

T3A. T3A adapts the classifier using target-subject context predictions. We initialize source prototypes from the source training set and update the prediction rule using only unlabeled context samples from the target subject. Query samples are used only for evaluation.

Table 7: Sensitivity analysis of the coordinate supervision weight λ_{coord} using CBraMod.

λ_{coord}	FACED		PhysioNet-MI		ISRUC	
	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑
0	60.89±0.82	55.92±0.91	65.59±1.13	54.07±1.16	81.15±0.94	76.42±1.44
0.01	61.99±0.85	57.14±1.02	65.60±0.92	54.08±1.04	82.16±0.88	77.11±1.20
0.03	61.90±0.77	55.08±0.84	66.10±0.91	54.77±1.09	81.60±0.93	76.95±1.01
0.1	62.63±0.74	57.90±0.75	67.21±0.82	56.24±1.07	82.87±0.81	78.32±1.24
0.3	62.26±0.79	57.32±0.83	66.86±0.98	55.78±1.09	82.69±0.82	78.61±1.26
1.0	59.83±0.90	54.42±0.99	65.23±1.06	53.60±1.23	81.11±0.87	76.93±1.32

Table 8: Sensitivity analysis of the canonical response preservation weight λ_{can} using CBraMod.

λ_{can}	FACED		PhysioNet-MI		ISRUC	
	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑	Bal. Acc. ↑	Kappa ↑
0	60.09±0.85	55.56±0.87	66.72±0.96	55.60±1.33	81.39±0.96	77.93±1.56
0.01	60.25±0.80	55.51±0.83	66.91±0.93	55.84±1.16	81.82±0.87	78.21±1.47
0.03	60.99±0.77	56.26±0.85	67.08±0.86	56.09±1.08	82.15±0.70	78.03±1.06
0.1	62.63±0.74	57.90±0.75	67.21±0.82	56.24±1.07	82.87±0.81	78.32±1.24
0.3	61.15±0.82	56.48±0.88	66.85±0.97	55.77±1.09	82.87±0.92	78.32±1.29
1.0	58.49±0.87	53.78±0.93	66.19±1.01	54.88±1.19	82.53±0.97	77.84±1.27

COME-TTA. COME-TTA performs test-time adaptation by optimizing an entropy-based objective with a consistency regularization mechanism on unlabeled target samples. Following the original setting, adaptation is conducted only on the target-subject context samples, while query samples are kept unseen during adaptation and used only for evaluation. Unlike SCC, which estimates a reusable subject coordinate and canonicalizes query representations without updating model parameters, COME-TTA adapts the model parameters through test-time optimization.

Foundation-model baselines. For foundation-model baselines, we use the same backbone wrappers, input preprocessing, classifier head, optimizer, and validation protocol as SCC. The only difference is that SCC additionally estimates and removes a subject coordinate from the query representation using the unlabeled context set.

E Additional Experimental Results

E.1 Hyperparameter Selection

We investigate the sensitivity of the two main SCC loss weights, λ_{coord} and λ_{can}, using CBraMod as the backbone. Each weight is varied independently while fixing the other to its default value. Table 7 evaluates the effect of λ_{coord}, which controls the supervision strength for subject-coordinate estimation. Removing this term ($\lambda_{\text{coord}} = 0$) consistently degrades performance, indicating that explicit coordinate supervision is important for disentangling subject-dependent displacement from task-related responses. Increasing the weight to moderate values improves unseen-subject decoding, with $\lambda_{\text{coord}} = 0.1$ providing a favorable balance across all three datasets. However, an excessively large weight ($\lambda_{\text{coord}} = 1.0$) causes performance degradation, suggesting that over-emphasizing coordinate alignment may restrict the representation adaptation and interfere with task discrimination. Table 8 studies the influence of λ_{can}, which encourages the canonical representation to preserve shared task-related responses. Without this constraint ($\lambda_{\text{can}} = 0$), performance decreases, particularly on FACED, demonstrating that removing subject coordinates alone is insufficient to guarantee preservation of discriminative geometry. A moderate response preservation weight improves performance, and $\lambda_{\text{can}} = 0.1$ achieves consistently strong results across different decoding tasks. When the weight becomes excessively large, the benefit diminishes, likely because overly strong response regularization limits the flexibility of subject canonicalization.

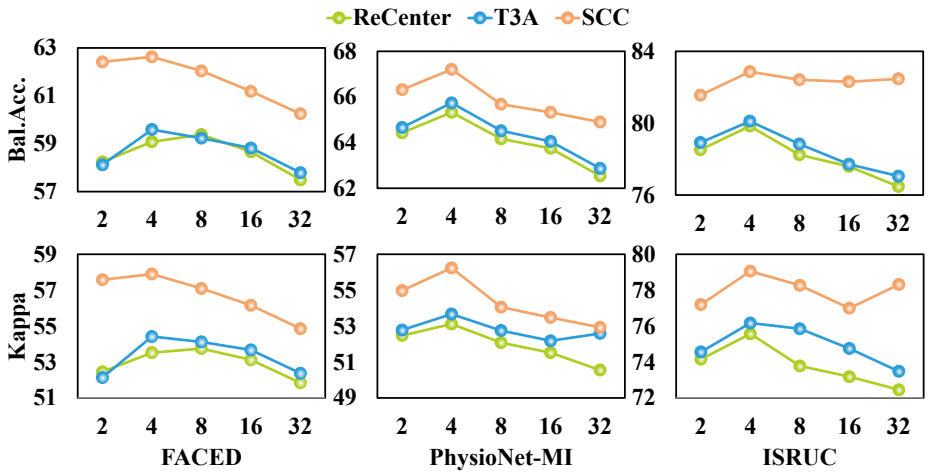

Figure 5: Additional context-size robustness results on CBraMod. We vary the number of unlabeled context samples per target subject at inference and report balanced accuracy (top) and Cohen's kappa (bottom) on FACED, PhysioNet-MI, and ISRUC. SCC consistently outperforms the compared context-aware baselines across datasets and achieves its best or near-best performance with a small context size, indicating effective adaptation in the few-shot unlabeled setting.

Table 9: Fixed-context nested evaluation of SCC using CBraMod. For each unseen subject, a maximum context pool of $M = 32$ unlabeled samples is first sampled, and smaller contexts are constructed as nested subsets of the same pool. The reported values are balanced accuracy. This protocol isolates the effect of context quantity by keeping context composition and query samples unchanged across different context sizes.

Dataset	$M = 2$	$M = 4$	$M = 8$	$M = 16$	$M = 32$
FACED	62.10	62.35	62.48	62.55	62.60
PhysioNet-MI	66.90	67.10	67.25	67.30	67.35
ISRUC	82.40	82.70	82.85	82.90	82.95

E.2 Additional Context Robustness Results

Figure 5 provides additional context-size robustness results on CBraMod. Under random sampling, SCC remains effective across different context sizes, with $M = 4$ achieving the best or near-best performance. Since random scaling changes both the number and composition of context samples, we further conduct a fixed-context nested evaluation in Appendix E.3. The results show that increasing context size consistently improves or maintains performance when context composition is controlled. These findings indicate that SCC benefits from informative subject context, while a small unlabeled calibration set is already sufficient in practical deployment scenarios.

E.3 Fixed-context Nested Context Evaluation

Table 9 examines whether the degradation observed under random large contexts is caused by the increased number of context samples or by changes in context composition. When context composition is controlled, increasing the context size from $M = 2$ to $M = 32$ consistently improves performance across FACED, PhysioNet-MI, and ISRUC. For example, FACED improves from 62.10% to 62.60%, while similar monotonic improvements are observed on PhysioNet-MI and ISRUC. These results indicate that averaging additional coordinate proposals is beneficial when the context distribution remains consistent. Therefore, the degradation observed in the original random context-size analysis is not caused by excessive context samples themselves. Instead, it is likely related to the changing composition of randomly sampled contexts, which may introduce different class distributions or within-subject variations. This finding further supports the design choice of SCC: effective

adaptation depends on obtaining representative subject-level context rather than simply increasing calibration quantity.