PACE: PROGRESSIVE AFFECTIVE ALIGNMENT FOR EEG FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG foundation models can encode useful affective information, but their representations are not necessarily organized for transfer across heterogeneous subjects, datasets, and elicitation contexts. We diagnose frozen representations from four EEG foundation models and find that emotion relations are substantially more stable within subjects than across heterogeneous datasets and elicitation contexts, revealing a gap between recoverable affective information and transferable affective organization. We introduce Progressive Affective Contrastive Enhancement (PACE), an affective post-training framework that progressively expands affective alignment from reliable within-dataset relations to heterogeneous datasets and elicitation contexts. PACE first learns cross-subject affective structure within individual datasets, then extends alignment across datasets and elicitation contexts while retaining within-dataset supervision. We evaluate PACE on twelve completely held-out target datasets and four foundation backbones. PACE improves average full-fine-tuning balanced accuracy by 2.66–7.05 percentage points and linear-probing performance by 3.52–5.77 points. Moreover, PACE outperforms CE, SupCon, and VREx post-training approaches on all twelve target datasets. Stage ablations support the progressive strategy, and frozen-representation analyses show improved emotion-relation geometry across all four backbones. Together, these results establish affective post-training as an effective bridge between generic EEG pretraining and transferable affective representations across heterogeneous datasets and elicitation contexts.Code is available at `https://anonymous.4open.science/r/ICLR2027_PACE/`.

1 INTRODUCTION

Large-scale EEG foundation models have opened new opportunities for learning generalizable neural representations (Jiang et al., 2024; Wang et al., 2025; Zhou et al., 2025; El Ouahidi et al., 2025), yet their ability to support generalization in emotion decoding remains unclear. Although pretrained models may contain emotion-related information, whether such information forms context-invariant representations remains unclear. EEG emotion decoding is particularly challenged by substantial inter-subject variability, dataset shifts, and differences in elicitation contexts and annotation schemes (Apicella et al., 2024; Chen et al., 2023; Xiao et al., 2024; Xu et al., 2025). Similar affective states can be induced by videos, sounds, or imagined experiences, expressed differently across individuals, and annotated under heterogeneous emotion taxonomies (Chen et al., 2023; Xiao et al., 2024; Xu et al., 2025). In fMRI studies, affective neural representations have been shown to exhibit both shared and stimulus-specific organization across different stimulus types (Čeko et al., 2022), motivating the search for transferable affective representations across contexts in EEG. An important open question is how to transform generic EEG representations into an affective geometry that remains stable across heterogeneous datasets and elicitation contexts.

Despite rapid progress in EEG foundation models, existing studies have primarily evaluated transferability through downstream adaptation on individual benchmark datasets (Jiang et al., 2024; Wang et al., 2025; Zhou et al., 2025; El Ouahidi et al., 2025; Kommineni et al., 2026). Whether these generic representations preserve affective structures that remain comparable across heterogeneous emotion datasets and elicitation contexts remains largely unexplored. As illustrated in Figure 1(a), we diagnose frozen representations from four EEG foundation backbones by measuring emotion relations across progressively broader boundaries: within subjects, across subjects, across datasets,

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.

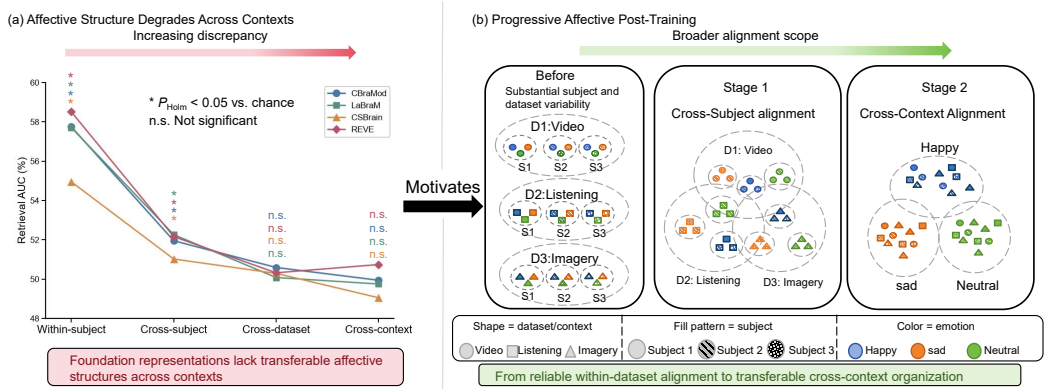

Figure 1: **Motivation and overview of PACE.** (a) Affective structure degrades across broader transfer boundaries. (b) PACE progressively expands alignment from cross-subject to cross-context settings.

and across contexts. We find that pretrained representations can partially recover emotion relations within-subject settings, yet this organization rapidly deteriorates when moving to cross-subject, cross-dataset, and cross-context settings, approaching chance levels in broader transfer boundaries. These observations reveal a fundamental gap between encoding affective information and organizing it into a representation space that generalizes across subjects, datasets, and elicitation contexts.

Recent efforts toward generalizable EEG emotion recognition have mainly addressed subject variability. Contrastive approaches such as CLISA exploit shared stimulus responses to learn subject-invariant representations by aligning neural responses across individuals (Shen et al., 2023; Chang et al., 2025). More recent multi-dataset learning methods further explore how to leverage heterogeneous emotion datasets through joint representation learning (Zhang et al., 2025; Chen et al., 2026). However, affective alignment across datasets and elicitation contexts remains challenging due to limited experimental correspondence and heterogeneous emotion taxonomies. Previous studies have shown that subject-level alignment can improve cross-dataset transfer even without explicit dataset correspondence (Zhang et al., 2025), suggesting that reliable subject-level organization may provide a foundation for broader affective generalization. Nevertheless, how to systematically learn transferable affective representations across subjects, datasets, and elicitation contexts, and how to fully exploit shared emotional structures across heterogeneous contexts, remain largely unexplored.

We introduce **Progressive Affective Contrastive Enhancement (PACE)**, an affective post-training framework that specializes EEG foundation models by progressively organizing transferable affective representations across subjects, datasets, and elicitation contexts (Figure 1b). Rather than directly aligning heterogeneous emotion datasets, PACE first establishes reliable affective organization within each source dataset through cross-subject correspondence using shared stimulus or temporal structures and native emotion labels (Stage 1). It then extends this organization to broader affective spaces by aligning compatible emotion relations across datasets and elicitation contexts while preserving within-dataset supervision (Stage 2). This progressive strategy enables EEG foundation models to learn transferable affective structures across heterogeneous contexts. We evaluate PACE under a strict leave-one-dataset-out setting across twelve held-out targets and four EEG foundation backbones, demonstrating consistent improvements in both downstream transfer and affective representation organization. Our contributions are three-fold:

1. **Revealing the gap between within-subject affective information and transferable organization.** We develop a unified relation-ranking framework to characterize frozen EEG foundation-model representations across progressively broader transfer settings and reveal that pretrained models can preserve within-subject affective structures while failing to organize emotion representations generalizable across heterogeneous datasets and elicitation contexts.
2. **A progressive framework for affective representation alignment.** We introduce PACE, which establishes reliable affective organization through within-dataset cross-subject align-

ment before extending it to heterogeneous datasets and elicitation contexts. By progressively expanding affective supervision while retaining reliable within-dataset relations, PACE provides a principled strategy for specializing generic EEG foundation models toward transferable affective representations.

3. **Comprehensive evaluation of transferable affective representations.** Across twelve held-out target datasets and four EEG foundation backbones, PACE consistently improves affective transfer, achieving average balanced-accuracy gains of 2.66–7.05 percentage points under full fine-tuning and 3.52–5.77 percentage points under linear probing. PACE further outperforms CE, SupCon, and VREx post-training approaches on all twelve CBraMod targets, with stage ablations and frozen-representation analyses confirming improved affective organization across heterogeneous transfer settings.

2 Related Work

2.1 EEG Foundation Models

Large-scale self-supervised pretraining has recently enabled EEG foundation models to learn general-purpose neural representations from heterogeneous recordings. LaBraM (Jiang et al., 2024) introduces masked neural-code prediction by discretizing EEG signals into neural tokens, while CBraMod (Wang et al., 2025) adopts masked EEG reconstruction with a criss-cross architecture to capture spatiotemporal dependencies. Other recent efforts, including REVE (El Ouahidi et al., 2025) and CSBrain (Zhou et al., 2025), further explore large-scale EEG pretraining through reconstruction-based objectives and cross-scale representation learning. These models establish powerful generic representations for diverse EEG decoding tasks, but their pretraining objectives are largely task-agnostic. Recent work explores finetuning REVE with supervised contrastive learning (SupCon) for cross-subject emotion recognition (Pasinato, 2026). However, whether these representations capture affective structures that remain transferable across heterogeneous emotion datasets and elicitation contexts remains unclear. This motivates our affective post-training framework, which further adapts pretrained EEG foundation models with affective supervision to improve the transferability of emotion representations without retraining from scratch.

2.2 Generalizable Affective Representation Learning from EEG

Generalization across subjects has long been a central challenge in EEG emotion recognition due to substantial inter-individual variability. Existing approaches mainly address this issue through domain adaptation or invariant representation learning. Domain adaptation methods reduce subject-related distribution shifts by aligning feature distributions across individuals (Li et al., 2020; Huang et al., 2025). For domain generalization, VREx encourages generalization by penalizing risk variance across source subjects Krueger et al. (2021). It achieves the best performance in comparison to Deep CORAL, GroupDRO, and DANN in subject-independent emotion recognition (Kim et al., 2025). DMMR (Wang et al., 2024) further improves cross-subject generalization through denoising mixed mutual reconstruction, and CLISA (Shen et al., 2023) exploits shared stimulus responses and contrastive alignment to learn subject-invariant representations. Beyond subject-level variability, distillation-based domain generalization approaches investigate cross-dataset EEG emotion recognition by transferring invariant knowledge from multiple source domains to unseen target datasets (Li et al., 2025a).

More recently, studies have explored leveraging heterogeneous affective EEG datasets through joint representation learning. mdJPT (Zhang et al., 2025) performs multi-dataset joint pretraining on emotional EEG resources to improve generalizable affective representations. EMOD (Chen et al., 2026) integrates heterogeneous emotion annotations through valence–arousal-guided contrastive learning. However, heterogeneous datasets often differ in elicitation paradigms, recording protocols, and emotion taxonomies, making direct alignment challenging. Existing approaches typically address subject and dataset variability separately, while systematic learning of transferable affective representations across subjects, datasets, and elicitation contexts remains largely unexplored.

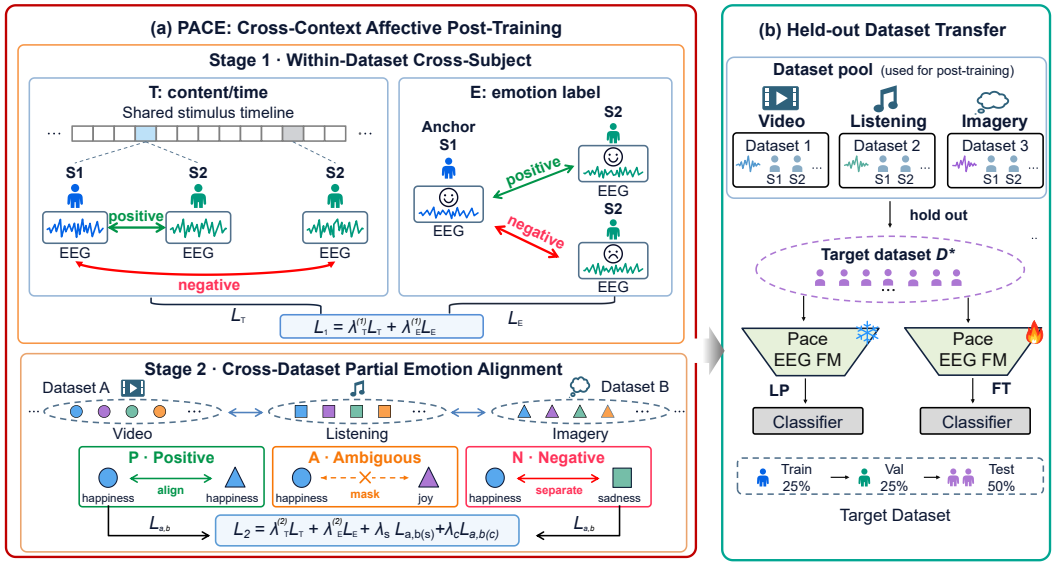

Figure 2: Overview of PACE. Stage 1 learns within-dataset cross-subject structure, Stage 2 aligns emotions across datasets and paradigms, and the resulting representation is transferred to a held-out target for linear probing (LP) or full finetuning (FT).

3 METHOD

Given multiple source emotion-EEG datasets, PACE applies a two-stage affective post-training to an EEG foundation encoder and transfers the resulting representation to a held-out target dataset (Figure 2). Stage 1 establishes cross-subject affective structure within each source dataset using stimulus or temporal correspondence and native emotion labels. Stage 2 extends alignment across datasets and elicitation contexts through compatible emotion relations, while retaining within-dataset supervision to preserve the structure learned in Stage 1.

3.1 PROBLEM FORMULATION AND SHARED REPRESENTATION

We consider M source emotion EEG datasets used for affective post-training,

$$\mathcal{D}_{\text{src}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_M\}, \qquad \mathcal{D}^\star \notin \mathcal{D}_{\text{src}}, \tag{1}$$

where $\mathcal{D}^\star$ denotes the target dataset that is completely held out during post-training. For an EEG segment x_i, we associate its dataset d_i, subject s_i, dataset-specific emotion label y_i, and, when available, stimulus/content c_i, temporal information t_i, and elicitation paradigm π_i.

Starting from a pretrained EEG foundation encoder f_θ, we map all source datasets into a shared representation space through a projection head g_ϕ:

$$z_i = \frac{g_\phi(f_\theta(x_i))}{\|g_\phi(f_\theta(x_i))\|_2}, \qquad sim_{ij} = z_i^\top z_j. \tag{2}$$

The same encoder and representation space are used throughout post-training.

3.2 WITHIN-DATASET CROSS-SUBJECT STRUCTURE LEARNING

Stage 1 operates within each dataset, where subjects share the same experimental design and emotion label space, allowing reliable cross-subject relations to be defined. We construct two complementary forms of supervision.

Table 1: Held-out transfer with FT. Mean balanced accuracy (%); bold marks gains over the backbone.

	Video								Listening		Imagery		
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg
Task-specific models													
EEGNet	51.57	29.34	25.19	20.69	37.36	26.30	22.20	52.88	32.95	61.53	14.41	8.67	31.92
TSception	47.52	29.36	25.88	18.84	19.39	19.25	25.51	51.08	33.83	64.13	14.33	8.55	29.81
LGGNet	51.87	32.22	25.60	19.90	30.58	21.20	24.82	51.92	31.09	60.41	14.13	7.81	30.96
DMMR	47.04	32.06	25.25	19.11	12.62	17.85	25.79	50.39	32.18	62.30	14.84	8.23	28.97
HEDN	45.27	32.55	23.50	18.87	11.09	16.38	20.79	49.83	30.08	62.01	14.11	7.50	27.67
CLISA	49.78	29.28	23.40	18.68	40.79	15.74	21.86	52.67	25.51	52.85	14.56	6.96	29.34
FreqDGT	46.08	30.64	23.22	17.65	14.48	15.60	23.82	50.08	28.77	58.34	14.90	7.23	27.57
Emotion-pretrained models													
mdJPT	52.55	33.59	25.76	20.51	19.56	19.57	23.72	51.61	31.36	63.05	13.73	7.36	30.20
MMM	49.58	30.12	24.68	18.05	15.38	15.38	22.16	50.76	33.17	57.30	14.76	6.84	28.18
EMOD	45.49	28.69	23.33	17.16	24.12	17.76	23.84	50.42	30.70	61.60	14.42	9.19	28.89
EEG foundation models													
CBraMod	53.35	36.38	29.68	23.99	33.45	20.32	26.17	52.45	36.21	61.01	13.98	8.71	32.98
+ PACE	**64.09**	**44.04**	**38.56**	**26.50**	**54.40**	**34.19**	**31.54**	**58.10**	**39.68**	**64.35**	**15.39**	**9.53**	**40.03**
Δ BAcc (pp)	**+10.74**	**+7.66**	**+8.88**	**+2.51**	**+20.95**	**+13.87**	**+5.37**	**+5.65**	**+3.47**	**+3.34**	**+1.41**	**+0.82**	**+7.05**
REVE	61.25	41.28	33.49	24.22	44.05	29.94	32.60	55.34	43.15	67.64	15.53	10.94	38.29
+ PACE	**69.30**	**46.03**	**42.81**	**29.54**	37.07	**34.88**	**37.65**	**56.55**	**44.02**	**68.23**	14.68	10.53	**40.94**
Δ BAcc (pp)	**+8.05**	**+4.75**	**+9.32**	**+5.32**	−6.98	**+4.94**	**+5.05**	**+1.21**	**+0.87**	**+0.59**	−0.85	−0.41	**+2.66**
CSBrain	53.29	35.13	29.73	22.56	33.03	19.29	27.26	52.22	35.22	61.76	14.63	8.34	32.70
+ PACE	**58.14**	**39.67**	**34.53**	**25.26**	**44.69**	**28.35**	**28.20**	**55.85**	**36.62**	**64.78**	**14.81**	**8.92**	**36.65**
Δ BAcc (pp)	**+4.85**	**+4.54**	**+4.80**	**+2.70**	**+11.66**	**+9.06**	**+0.94**	**+3.63**	**+1.40**	**+3.02**	**+0.18**	**+0.58**	**+3.95**
LaBraM	60.08	39.65	32.77	23.34	37.54	18.93	29.11	54.06	38.20	65.73	15.67	9.67	35.40
+ PACE	**65.21**	**41.88**	**37.80**	**27.58**	**38.60**	**26.59**	**32.81**	**56.73**	**41.31**	**67.21**	14.47	9.19	**38.28**
Δ BAcc (pp)	**+5.13**	**+2.23**	**+5.03**	**+4.24**	**+1.06**	**+7.66**	**+3.70**	**+2.67**	**+3.11**	**+1.48**	−1.20	−0.48	**+2.89**

The first relation, denoted T, captures stimulus or temporal correspondence across subjects. When two samples originate from different subjects but correspond to the same stimulus, content, or temporal segment, they are treated as positive pairs across subjects:

$$(i, j) \in P_i^{(T)} \iff d_i = d_j,\ s_i \neq s_j,\ (c_i, t_i) \sim (c_j, t_j). \tag{3}$$

The second relation, denoted E, exploits the native emotion labels of each dataset. For cross-subject samples within the same dataset, positive pairs ($P_i^{(E)}$) and negative pairs ($N_i^{(E)}$) are defined as:

$$\begin{aligned} (i, j) \in P_i^{(E)} &\iff d_i = d_j,\ s_i \neq s_j,\ y_i = y_j, \\ (i, j) \in N_i^{(E)} &\iff d_i = d_j,\ s_i \neq s_j,\ y_i \neq y_j. \end{aligned} \tag{4}$$

These relations use each dataset's native emotion taxonomy, making within-dataset supervision available without a global label space.

For a relation r, with positive and negative sets $P_i^{(r)}$ and $N_i^{(r)}$, we use the multi-positive form of supervised contrastive learning (Khosla et al., 2020), with positives and negatives defined by the relation:

$$\mathcal{L}_r = -\frac{1}{|\mathcal{I}_r|} \sum_{i \in \mathcal{I}_r} \frac{1}{|P_i^{(r)}|} \sum_{j \in P_i^{(r)}} \log \frac{\exp(sim_{ij}/\tau)}{\sum_{k \in P_i^{(r)} \cup N_i^{(r)}} \exp(sim_{ik}/\tau)}, \tag{5}$$

where τ is the temperature and $\mathcal{I}_r$ contains anchors with valid positive relations.

The Stage-1 objective combines the two sources of within-dataset supervision:

$$\mathcal{L}_1 = \lambda_T^{(1)} \mathcal{L}_T + \lambda_E^{(1)} \mathcal{L}_E. \tag{6}$$

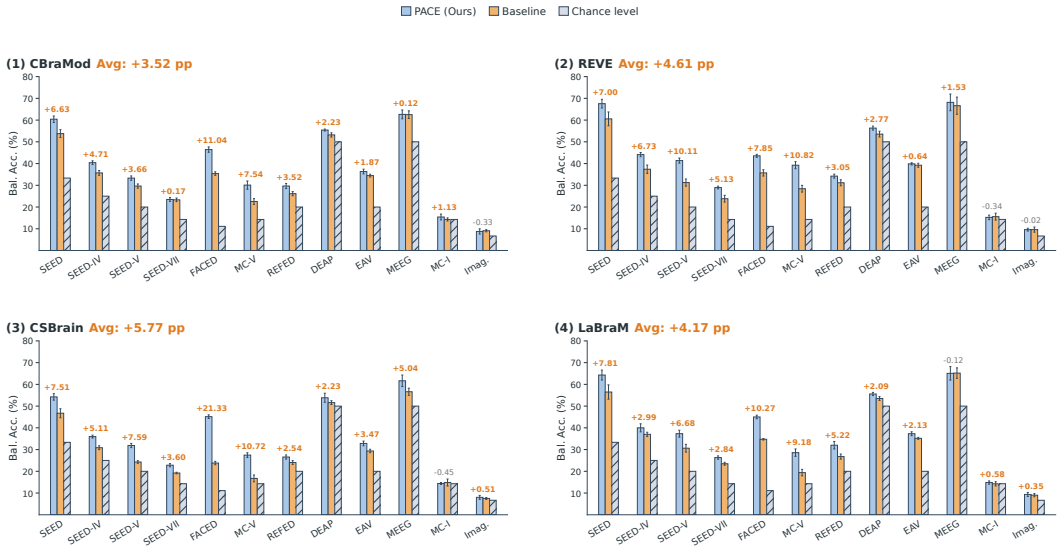

Figure 3: Linear probing across four backbones. Bars show balanced accuracy; annotations give PACE gains in pp and Avg averages all twelve targets. Error bars show standard deviations across six random seeds; hatched bars indicate chance.

Here, $\lambda_T^{(1)}$ and $\lambda_E^{(1)}$ weight the two Stage 1 losses; their values are given in Appendix B.1. The T provides reliable cross-subject pairs based on shared stimulus or temporal information, whereas E provides additional pairs based on native emotion labels. Together, they establish cross-subject affective structure before cross-dataset heterogeneity is introduced.

3.3 Cross-Dataset Emotion Alignment

Stage 2 extends the affective structure learned within individual datasets to alignment across datasets and elicitation contexts. Because datasets use different emotion taxonomies, their native labels cannot be directly compared. We therefore align samples whose labels express compatible emotion concepts, while retaining the within-dataset objectives to preserve the structure learned in Stage 1.

Specifically, we use a predefined mapping h_d to map each dataset's native labels to shared emotion concepts in $\mathcal{V}$, assigning $\perp$ to labels without a suitable match. Across datasets, samples mapped to the same concept form positive pairs, whereas those mapped to different concepts form negative pairs. To reduce potentially unreliable negative supervision, we exclude a predefined set of concept pairs whose distinction may be ambiguous across datasets, such as joy and happiness. Such ambiguous concept pairs are specified by a fixed exclusion set $\mathcal{B}$. Appendix B.4 provides the complete specification.

For each dataset pair (a, b), we construct contrastive pairs between samples from the two datasets. Let $\ell_i = h_{d_i}(y_i)$ be the mapped emotion concept of sample i. For an anchor i from either dataset, its positive and negative sets contain samples from the other dataset:

$$\begin{aligned} P_i^{a,b} &= \{j : d_j \in \{a,b\},\ d_j \neq d_i,\ \ell_i = \ell_j \in \mathcal{V}\}, \\ N_i^{a,b} &= \{j : d_j \in \{a,b\},\ d_j \neq d_i,\ \ell_i, \ell_j \in \mathcal{V},\ \ell_i \neq \ell_j,\ \{\ell_i, \ell_j\} \notin \mathcal{B}\}. \end{aligned} \tag{7}$$

Pairs involving unmapped labels or excluded concept combinations are omitted from the cross-dataset loss.

The contrastive loss for dataset pair (a, b) is

$$\mathcal{L}_{a,b} = -\frac{1}{|\mathcal{I}_{a,b}|} \sum_{i \in \mathcal{I}_{a,b}} \frac{1}{|P_i^{a,b}|} \sum_{j \in P_i^{a,b}} \log \frac{\exp(sim_{ij}/\tau)}{\sum_{k \in P_i^{a,b} \cup N_i^{a,b}} \exp(sim_{ik}/\tau)}, \tag{8}$$

where $\mathcal{I}_{a,b}$ contains anchors from datasets a and b with at least one valid positive. This objective increases similarity between emotion-matched samples relative to eligible negatives from the other dataset.

We group eligible dataset pairs into G_s and G_c, corresponding to the same and different elicitation paradigms, respectively. Stage 2 combines the average cross-dataset loss within each group with the retained within-dataset objectives:

$$\mathcal{L}_2 = \underbrace{\lambda_T^{(2)}\mathcal{L}_T + \lambda_E^{(2)}\mathcal{L}_E}_{\text{within-dataset structure}} + \underbrace{\frac{\lambda_s}{|G_s|}\sum_{(a,b)\in G_s}\mathcal{L}_{a,b}}_{\text{same-paradigm alignment}} + \underbrace{\frac{\lambda_c}{|G_c|}\sum_{(a,b)\in G_c}\mathcal{L}_{a,b}}_{\text{cross-paradigm alignment}}. \tag{9}$$

The retained $\mathcal{L}_T$ and $\mathcal{L}_E$ use native emotion labels, while both cross-dataset groups use the shared concept mapping and the loss defined above. Only dataset pairs with valid positives are included. Appendix B.1 specifies the loss weights.

3.4 Cross-Context Affective Diagnosis

We also examine whether emotion-matched trials remain close in the representation space as comparisons extend across subjects, datasets, and elicitation contexts. By evaluating frozen representations directly, it assesses affective organization without fitting a downstream classifier. We compare frozen trial-level representations from each backbone and PACE using target-specific held-out checkpoints. Four settings examine progressively broader comparison boundaries: within-subject (WS) compares different trials of one subject; cross-subject (CS) compares trials from different subjects within a dataset; cross-dataset (CD) compares trials from different datasets within an elicitation paradigm; and cross-paradigm (CP) compares trials from datasets using different paradigms or contexts.

For each query q and an eligible negative emotion class c, let $\mathcal{P}_q$ contain eligible emotion-matched trials and $\mathcal{N}_{q,c}$ contain eligible trials from class c. Eligibility follows the comparison boundary and emotion-label rules of each setting. We measure how often an emotion-matched trial is more similar to the query than a negative trial. Using cosine similarity s, we define $\Delta_{qpn} = s(q,p) - s(q,n)$ and compute relation AUC through its pairwise-ranking interpretation (Hanley & McNeil, 1982):

$$\mathrm{AUC}(q,c) = \frac{1}{|\mathcal{P}_q||\mathcal{N}_{q,c}|}\sum_{p\in\mathcal{P}_q}\sum_{n\in\mathcal{N}_{q,c}}\left[\mathbf{1}\{\Delta_{qpn} > \epsilon\} + \frac{1}{2}\mathbf{1}\{|\Delta_{qpn}| \le \epsilon\}\right], \tag{10}$$

where $\epsilon = 10^{-7}$ treats numerical ties as half credit. We average scores over eligible negative classes for each query, then over queries within each target, and finally macro-average across eligible target datasets. Scores are reported as $100 \times \mathrm{AUC}$, with a chance level of 50. The backbone and PACE share evaluation samples and a fixed scoring rule, excluding identical trials and identifiable shared stimuli. Appendix E details feature extraction, sampling, relation eligibility, aggregation, and uncertainty estimation.

4 Experimental Setup

4.1 Datasets

We evaluate affective transfer on twelve emotion EEG datasets grouped into video, listening, and imagery settings. Video datasets comprise SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2019), SEED-V (Liu et al., 2022), SEED-VII (Jiang et al., 2025), FACED (Chen et al., 2023), the video subset of EmoEEG-MC (MC-V) (Xu et al., 2025), REFED (Ning et al., 2025), and DEAP (Koelstra et al., 2012). The listening group comprises EAV dataset's conversational listening condition (prerecorded videos) (Lee et al., 2024) and music-based MEEG dataset (Xiao et al., 2024). Imagery datasets comprise the imagery subset of EmoEEG-MC (MC-I) and ImaginedEmotion (Imag.) (Onton & Makeig, 2009). These datasets vary in subject populations, acquisition settings, stimulus content, and emotion taxonomies. Native labels define within-dataset supervision and downstream classification, while a partial shared vocabulary determines eligible cross-dataset relations. We applied a unified preprocessing pipeline to all datasets. See Appendix A for details of the datasets and preprocessing.

4.2 Leave-One-Dataset-Out Evaluation and Compared Models

In each round, one target dataset is excluded from affective post-training, and eligible non-target datasets supply PACE's training data. The target's EEG recordings, emotion labels, and pairwise relations are excluded from this post-training. MC-V and MC-I share subjects (Xu et al., 2025), so both subsets are excluded from the source pool when either is the target. We repeat the evaluation with each of the twelve datasets as the target, then adapt the transferred representation using that target's labeled training split. For each target dataset, we randomly partition subjects into disjoint training, validation, and test sets containing 25%, 25%, and 50% of the subjects, respectively. All recordings from a subject remain in the same partition. We repeat this subject-wise random split and evaluation with six random seeds. Full fine-tuning (FT) is the primary evaluation protocol, updating the transferred representation and classification head on the target training set. Linear probing (LP) is the secondary protocol, freezing the transferred representation and training a linear classifier on the same target split. Main tables show mean results of the six seeds. Appendix C reports complete values and standard deviations across the six seeds. We evaluate four foundation backbones and their PACE counterparts alongside EEG emotion-recognition and representation-learning baselines. Appendix B.3 gives the original model references.

5 Results

5.1 Held-Out Dataset Transfer

PACE improves the mean full-fine-tuning performance of all four EEG foundation backbones across twelve held-out targets (Table 1). The average performance gains are 7.05 percentage points (pp) for CBraMod, 2.66 pp for REVE, 3.95 pp for CSBrain, and 2.89 pp for LaBraM. Improvements occur in 43 of the 48 backbone–target comparisons, including all twelve targets for CBraMod and CSBrain. The improvement is observed across all four foundation backbones, indicating that the effect is not specific to a single pretrained architecture.

Linear probing results shows that PACE also improves the frozen representations. LP improvements average 3.52, 4.61, 5.77, and 4.17 pp for CBraMod, REVE, CSBrain, and LaBraM, respectively (Figure 3). PACE improves 43 of 48 comparisons and all 32 video-target comparisons. Improvements under both LP and FT indicate that PACE benefits the frozen representation itself as well as subsequent target-specific adaptation. Appendix C reports the complete balanced accuracy and macro-F1 results for FT and LP.

5.2 Staged Learning Ablation

We compare the pretrained backbone, the Stage 1-only and Stage 2-only variants, and the complete Stage 1→Stage 2 sequence on CBraMod (Table 2). The complete sequence reaches an average performance of 40.03% with FT, compared with 32.40% for the Stage 2-only variant, and improves every target in this comparison. This 7.63 pp gap shows that cross-dataset alignment alone is insufficient and that Stage 1 provides an important initialization for Stage 2. Stage 2 also adds transfer gains to the Stage 1 representation. Across the twelve FT targets, PACE improves on Stage 1 in eleven cases, with a mean paired gain of 1.93 pp. The largest additions occur on FACED (+5.17 pp), SEED (+5.11 pp), and SEED-IV (+3.66 pp). LP shows the same progression in average performance. Complete FT and LP results with split standard deviations are provided in Appendix D.

Table 2: Stage ablation with CBraMod under FT. Mean balanced accuracy (%); best in bold, second best underlined.

	Video								Listening		Imagery		
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg
Backbone	53.35	36.38	29.68	23.99	33.45	20.32	26.17	52.45	36.21	61.01	13.98	8.71	32.98
S1 only	<u>58.98</u>	<u>40.38</u>	<u>35.21</u>	<u>25.89</u>	<u>49.23</u>	<u>32.59</u>	<u>30.58</u>	<u>56.00</u>	<u>39.38</u>	<u>64.15</u>	<u>15.25</u>	**9.59**	<u>38.10</u>
S2 only	50.60	33.02	28.43	21.26	41.78	19.45	24.43	52.14	34.75	59.64	14.00	9.25	32.40
PACE	**64.09**	**44.04**	**38.56**	**26.50**	**54.40**	**34.19**	**31.54**	**58.10**	**39.68**	**64.35**	**15.39**	<u>9.53</u>	**40.03**

5.3 Comparison with Alternative Post-Training Methods

We compare PACE with three complementary post-training baselines on the same CBraMod backbone and held-out target splits (Table 3): CE for direct emotion supervision using dataset-native labels, SupCon (Pasinato, 2026) for label-based contrastive representation alignment, and VREx (Krueger et al., 2021) for domain generalization with cross-subject risk regularization.

Table 3: Post-training methods with CBraMod under FT. Mean balanced accuracy (%); best in bold, second best underlined.

	Video								Listening		Imagery		
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg
Backbone	53.35	36.38	29.68	<u>23.99</u>	33.45	20.32	26.17	52.45	36.21	61.01	13.98	8.71	32.98
CE	<u>58.11</u>	37.51	31.73	23.50	33.99	20.59	<u>27.91</u>	53.18	<u>38.74</u>	63.03	14.86	<u>9.22</u>	34.36
SupCon	56.93	<u>37.61</u>	<u>31.83</u>	23.95	36.88	25.41	27.02	<u>53.21</u>	36.77	<u>63.26</u>	<u>15.04</u>	9.09	<u>34.75</u>
VREx	55.94	36.28	31.27	23.42	<u>38.76</u>	<u>25.63</u>	26.70	53.12	37.47	62.40	14.80	8.63	34.54
PACE	**64.09**	**44.04**	**38.56**	**26.50**	**54.40**	**34.19**	**31.54**	**58.10**	**39.68**	**64.35**	**15.39**	**9.53**	**40.03**

5.4 Affective Representation Analysis

We examine whether PACE improves affective organization in backbone representations through relation ranking and visualization. PACE significantly improves relation AUC in all sixteen backbone–setting comparisons across the four settings (within-subject, cross-subject, cross-dataset, and cross-paradigm), with all paired 95% bootstrap intervals excluding zero. Cross-paradigm gains range from 2.74 to 3.72 percentage points, indicating improved emotion-relation ranking across elicitation contexts. Consistent with these results, t-SNE visualizations (van der Maaten & Hinton, 2008) on SEED (Figure 4) show clearer emotion-specific neighborhoods across all four backbones, accompanied by kNN@10 gains of 20.08–27.63 percentage points. Appendices E and F provide detailed results.

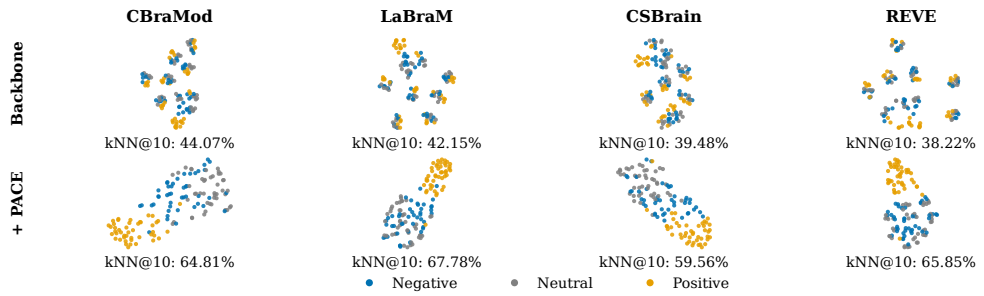

Figure 4: t-SNE on SEED before and after PACE. Colors denote emotions.

6 Conclusion

EEG foundation models encode affective information, but their representations do not readily generalize across subjects, datasets, and elicitation contexts. PACE addresses this gap by first establishing cross-subject affective structure within datasets, then extending alignment across heterogeneous contexts while retaining within-dataset supervision. Across twelve held-out targets and four EEG foundation backbones, PACE improves average performance under both linear probing and full fine-tuning, alongside consistent gains in frozen-representation emotion-relation ranking. Stage ablations further suggest that reliable within-dataset structure provides an important foundation for broader alignment. These findings demonstrate the value of progressive affective post-training for learning transferable emotion representations.

AI USE STATEMENT

We used generative AI tools for English-language polishing and LaTeX formatting. Their use was limited to improving the wording, readability, and presentation of the manuscript.

The authors take full responsibility for the final content of this work, including the methods, results, conclusions, and AI-assisted text and formatting.

REFERENCES

Andrea Apicella, Pasquale Arpaia, Giovanni D'Errico, Davide Marocco, Giovanna Mastrati, Nicola Moccaldi, and Roberto Prevete. Toward cross-subject and cross-session generalization in eeg-based emotion recognition: Systematic review, taxonomy, and methods. *Neurocomputing*, 604: 128354, 2024. doi: 10.1016/j.neucom.2024.128354.

Marta Čeko, Philip A. Kragel, Choong-Wan Woo, Marina López-Solà, and Tor D. Wager. Common and stimulus-type-specific brain representations of negative affect. *Nature Neuroscience*, 25(6): 760–770, 2022. doi: 10.1038/s41593-022-01082-w. URL https://www.nature.com/articles/s41593-022-01082-w.

Jiang Chang, Zhixin Zhang, Yuhua Qian, and Pan Lin. Multi-scale hyperbolic contrastive learning for cross-subject eeg emotion recognition. *IEEE Transactions on Affective Computing*, 16(3): 1716–1731, 2025.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A Large Finer-grained Affective Computing EEG Dataset. *Scientific Data*, 10:740, 2023. doi: 10.1038/s41597-023-02650-w. URL https://www.nature.com/articles/s41597-023-02650-w.

Yuning Chen, Sha Zhao, Shijian Li, and Gang Pan. EMOD: A Unified EEG Emotion Representation Framework Leveraging V-A Guided Contrastive Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17427–17435, 2026. doi: 10.1609/aaai.v40i21.38796. URL https://ojs.aaai.org/index.php/AAAI/article/view/38796.

Yi Ding, Neethu Robinson, Chengxuan Tong, Qiuhao Zeng, and Cuntai Guan. LGGNet: Learning From Local-Global-Graph Representations for Brain–Computer Interface. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7):9773–9786, 2024. doi: 10.1109/TNNLS.2023.3236635. URL https://doi.org/10.1109/TNNLS.2023.3236635.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A Foundation Model for EEG - Adapting to Any Setup with Large-Scale Pretraining on 25,000 Subjects. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0760. URL https://papers.neurips.cc/paper_files/paper/2025/hash/20a917f77773ac0fa8bea2bdd6606b66-Abstract-Conference.html.

J A Hanley and B J McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982. doi: 10.1148/radiology.143.1.7063747. URL https://pubmed.ncbi.nlm.nih.gov/7063747/.

He Huang, Xiaopeng Si, Yumeng Han, and Dong Ming. A novel conditional adversarial domain adaptation network for eeg cross-subject emotion recognition. *IEEE Transactions on Affective Computing*, 16(4):2905–2917, 2025.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/47393e8594c82ce8fd83adc672cf9872-Abstract-Conference.html.

Wei-Bang Jiang, Xuan-Hao Liu, Wei-Long Zheng, and Bao-Liang Lu. SEED-VII: A Multimodal Dataset of Six Basic Emotions With Continuous Labels for Emotion Recognition. *IEEE Transactions on Affective Computing*, 16(2):969–985, 2025. doi: 10.1109/TAFFC.2024.3485057. URL https://doi.org/10.1109/TAFFC.2024.3485057.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised Contrastive Learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html.

Jung-Hwan Kim, Hyerin Nam, Doyeon Won, and Chang-Hwan Im. Domain-generalized deep learning for improved subject-independent emotion recognition based on electroencephalography. *Experimental Neurobiology*, 34(3):119–130, 2025. doi: 10.5607/en25011.

Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012. doi: 10.1109/T-AFFC.2011.15. URL https://doi.org/10.1109/T-AFFC.2011.15.

Aditya Kommineni, Emily Zhou, Kleanthis Avramidis, Tiantian Feng, and Shrikanth Narayanan. A multi-dimensional framework for evaluating generalization in eeg foundation models. In *Proceedings of the 7th Conference on Health, Inference, and Learning*, volume 333 of *Proceedings of Machine Learning Research*, pp. 221–244. PMLR, 2026. URL https://proceedings.mlr.press/v333/kommineni26a.html.

David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, and Aaron Courville. Out-of-Distribution Generalization via Risk Extrapolation (REx). In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 5815–5826, 2021. URL https://proceedings.mlr.press/v139/krueger21a.html.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c. URL https://doi.org/10.1088/1741-2552/aace8c.

Min-Ho Lee, Adai Shomanov, Balgyn Begim, Zhuldyz Kabidenova, Aruna Nyssanbay, Adnan Yazici, and Seong-Whan Lee. EAV: EEG-Audio-Video Dataset for Emotion Recognition in Conversational Contexts. *Scientific Data*, 11:1026, 2024. doi: 10.1038/s41597-024-03838-4. URL https://www.nature.com/articles/s41597-024-03838-4.

Jinpeng Li, Shuang Qiu, Changde Du, Yixin Wang, and Huiguang He. Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*, 12(2):344–353, 2020.

Wei Li, Siyi Wang, Shitong Shao, and Kaizhu Huang. Distillation-based domain generalization for cross-dataset eeg-based emotion recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(3):2474–2490, 2025a.

Yueyang Li, Shengyu Gong, Weiming Zeng, Nizhuan Wang, and Wai Ting Siok. FreqDGT: Frequency-Adaptive Dynamic Graph Networks with Transformer for Cross-Subject EEG Emotion Recognition. In *2025 International Conference on Machine Intelligence and Nature-Inspired Computing (MIND)*, pp. 21–26, 2025b. doi: 10.1109/MIND67540.2025.11351599. URL https://doi.org/10.1109/MIND67540.2025.11351599.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2):715–729, 2022. doi: 10.1109/TCDS.2021.3071170. URL https://doi.org/10.1109/TCDS.2021.3071170.

Xiaojun Ning, Jing Wang, Zhiyang Feng, Tianzuo Xin, Shuo Zhang, Shaoqi Zhang, Zheng Lian, Yi Ding, Youfang Lin, and Ziyu Jia. REFED: A Subject Real-time Dynamic Labeled EEG-fNIRS Synchronized Recorded Emotion Dataset. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.neurips.cc/paper_files/paper/2025/file/2bf0ed7c35d9d84128e7f7c72ab76402-Paper-Datasets_and_Benchmarks_Track.pdf.

Julie Onton and Scott Makeig. High-frequency broadband modulations of electroencephalographic spectra. *Frontiers in Human Neuroscience*, 3:61, 2009. doi: 10.3389/neuro.09.061.2009. URL https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/neuro.09.061.2009/full.

Alberto Pasinato. Supervised contrastive fine-tuning of EEG foundation models for cross-subject emotion recognition. Master's thesis, Delft University of Technology, 2026. URL https://resolver.tudelft.nl/uuid:0af8c7ee-7478-48eb-a8b2-e18dfdb7cf73.

Xinke Shen, Xianggen Liu, Xin Hu, Dan Zhang, and Sen Song. Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition. *IEEE Transactions on Affective Computing*, 14(3):2496–2511, 2023. doi: 10.1109/TAFFC.2022.3164516. URL https://doi.org/10.1109/TAFFC.2022.3164516.

Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL https://www.jmlr.org/papers/v9/vandermaaten08a.html.

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 *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/bbbd6d915cb90be21c1254a82d45cedd-Abstract-Conference.html.

Yiming Wang, Bin Zhang, and Yujiao Tang. DMMR: Cross-Subject Domain Generalization for EEG-Based Emotion Recognition via Denoising Mixed Mutual Reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 628–636, 2024. doi: 10.1609/aaai.v38i1.27819. URL https://ojs.aaai.org/index.php/AAAI/article/view/27819.

Minghao Xiao, Zhengxi Zhu, Kang Xie, and Bin Jiang. MEEG and AT-DGNN: Improving EEG Emotion Recognition with Music Introducing and Graph-based Learning. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 4201–4208, 2024. doi: 10.1109/BIBM62325.2024.10821806. URL https://ieeexplore.ieee.org/document/10821806/.

Xin Xu, Xinke Shen, Xuyang Chen, Qingzhu Zhang, Sitian Wang, Yihan Li, Zongsheng Li, Dan Zhang, Mingming Zhang, and Quanying Liu. A Multi-Context Emotional EEG Dataset for Cross-Context Emotion Decoding. *Scientific Data*, 12:1142, 2025. doi: 10.1038/s41597-025-05349-2. URL https://www.nature.com/articles/s41597-025-05349-2.

Ke Yi, Yansen Wang, Kan Ren, and Dongsheng Li. Learning Topology-Agnostic EEG Representations with Geometry-Aware Modeling. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2344. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/a8c893712cb7858e49631fb03c941f8d-Abstract-Conference.html.

Qingzhu Zhang, Jiani Zhong, Zongsheng Li, Xinke Shen, and Quanying Liu. Multi-dataset Joint Pre-training of Emotional EEG Enables Generalizable Affective Computing. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5515. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/f1b8d443042f376aa3654b6c68de6297-Abstract-Conference.html.

Wei-Long Zheng and Bao-Liang Lu. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015. doi: 10.1109/TAMD.2015.2431497. URL https://doi.org/10.1109/TAMD.2015.2431497.

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics*, 49(3):1110–1122, 2019. doi: 10.1109/TCYB.2018.2797176. URL https://doi.org/10.1109/TCYB.2018.2797176.

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, 2025. doi: 10.52202/085713-2915. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/7e199ad8ae40eb19b2980f61f659cb07-Abstract-Conference.html.

ICLR 2027 — Supplementary Material

PACE: PROGRESSIVE AFFECTIVE ALIGNMENT FOR EEG FOUNDATION MODELS

A Datasets and Preprocessing

The twelve target datasets cover video, listening, and imagery paradigms. A shared preprocessing pipeline represents each recording as 10-s EEG segments with 60 channels sampled at 200 Hz.

A.1 Datasets and Evaluation Splits

Table 4 summarizes the subject counts, segment counts, and numbers of classification labels in the evaluated releases. Table 5 lists the corresponding label names. These labels define the downstream classification tasks: DEAP and MEEG use binary valence classification, while the remaining datasets use their categorical emotion labels. Each segment retains its trial label and subject/session identity throughout preprocessing. The MC counts refer to the evaluated subsets represented by the metadata and subject splits released with the code.

Table 4: Datasets in the unified 10-s release. Labels denotes the number of classes in the downstream classification task.

Dataset	Paradigm	Subjects	10-s rows	Labels
SEED	Video	15	14,895	3
SEED-IV	Video	15	14,655	4
SEED-V	Video	16	11,424	5
SEED-VII	Video	20	27,340	7
FACED	Video	123	10,332	9
MC-vid	Video	43	2,475	7
REFED	Video	32	4,800	5
DEAP	Video	32	7,680	2
EAV	Listening	41	8,200	5
MEEG	Listening	32	3,660	2
MC-ima	Imagery	42	2,457	7
ImaginedEmotion	Imagery	34	6,736	15

Table 5: Downstream label names in the evaluated releases. Names are listed in label-ID order, starting from zero. For REFED, L/M/H denote low/medium/high, and V/A denote valence/arousal.

Dataset	Label names
SEED	negative, neutral, positive
SEED-IV	neutral, sad, fear, happy
SEED-V	disgust, fear, sad, neutral, happy
SEED-VII	disgust, fear, sad, neutral, happy, anger, surprise
FACED	anger, disgust, fear, sadness, neutral, amusement, inspiration, joy, tenderness
MC-vid	joy, inspiration, tenderness, fear, disgust, sadness, neutral
REFED	MVMA (Neutral), LVLA (Sad), LVHA (Fear), HVHA (Happy), HVLA (Relax)
DEAP	low valence, high valence
EAV	Neutral, Sadness, Anger, Happiness, Calmness
MEEG	low valence, high valence
MC-ima	joy, inspiration, tenderness, fear, disgust, sadness, neutral
ImaginedEmotion	anger, awe, compassion, content, disgust, excite, fear, frustration, grief, happy, jealousy, joy, love, relief, sad

We evaluate transfer to an unseen EEG dataset with a known downstream emotion taxonomy. In each fold, affective post-training uses EEG recordings and sample-level annotations only from eligible source datasets, and all training pairs contain source samples. The target dataset is held out at the sample level throughout post-training. The shared concept mapping and relation matrix provide a fixed semantic prior over emotion category definitions (Appendix B.4). MC-vid and MC-ima share

subjects (Xu et al., 2025), so both subsets are excluded from the source pool when either is the target. Sources may share the target's elicitation paradigm or use a different paradigm. We repeat the evaluation with each of the twelve datasets as the target, then adapt the transferred representation using that target's labeled training split.

For each target dataset, subjects are randomly assigned to disjoint training, validation, and test sets in approximately 25%/25%/50% proportions. All recordings from a subject remain in the same partition. We repeat this subject-wise random split and evaluation with six random seeds. Full fine-tuning (FT) is the primary evaluation protocol, updating the transferred representation and classification head on the target training set. Linear probing (LP) is the secondary protocol, freezing the transferred representation and training a linear classifier on the same target split. Main tables show means; Appendix C reports complete values and split standard deviations across the six seeds.

A.2 Unified Preprocessing

Signals are converted to microvolts, filtered with a mains notch at 50 or 60 Hz and a zero-phase band-pass, and resampled to 200 Hz. The nominal passband is 0.3–75 Hz, subject to the source Nyquist limit. Filtering and resampling operate on continuous task recordings before segmentation. For a source signal X^{src}, the temporal processing is written as

$$U = \mathcal{R}_{200}(\mathcal{F}_{\mathrm{bp}}(\mathcal{F}_{\mathrm{notch}}(aX^{\mathrm{src}}))), \tag{11}$$

where a converts the source units to microvolts, $\mathcal{F}_{\mathrm{notch}}$ and $\mathcal{F}_{\mathrm{bp}}$ denote the notch and band-pass filters, and $\mathcal{R}_{200}$ denotes resampling. The resulting signal is divided into non-overlapping windows of 2,000 samples, with a stride of 2,000 samples.

A final remainder is retained if it contains at least 1,600 valid samples (8 s); shorter remainders are discarded. For a retained window starting at sample s_i with valid length ℓ_i, its canonical representation is

$$X_i(:,t) = \begin{cases} U(:, s_i + t), & 0 \le t < \ell_i, \\ 0, & \ell_i \le t < 2000, \end{cases} \qquad 1600 \le \ell_i \le 2000. \tag{12}$$

Thus, incomplete windows are padded with zeros only on the right, and their valid lengths are retained. Windows never cross trial, recording, session, or subject boundaries. Subsequent channel processing preserves these temporal boundaries and excludes padded samples when detecting bad channels.

Our single-pass bad-channel repair procedure is adapted from the amplitude-based detection and neighboring-channel interpolation strategy of mdJPT (Zhang et al., 2025). Detection uses the retained valid samples within each trial, with one shared bad-channel decision for all windows from that trial. Only scalp EEG channels with valid montage coordinates are eligible; auxiliary and reference channels are excluded. Let $\mathcal{C}$ denote the eligible channels and $\mathcal{V}$ the valid sample indices within a trial. We define

$$\begin{aligned} m &= \mathrm{median}_{c\in\mathcal{C},\, t\in\mathcal{V}}\, |x_c(t)|, \\ \rho_c(b) &= \frac{1}{|\mathcal{V}|} \sum_{t\in\mathcal{V}} \mathbf{1}[|x_c(t)| > b]\,. \end{aligned} \tag{13}$$

A channel is marked bad if any of four conditions holds: $\rho_c(3m) > 0.40$, $\rho_c(30m) > 0.01$, $\mathrm{std}_{t\in\mathcal{V}}\, x_c(t) < 0.001\,\mu\mathrm{V}$, or $\rho_c(800\,\mu\mathrm{V}) > 0.10$. These criteria identify sustained relative-amplitude artifacts, brief bursts, flat signals, and extreme absolute amplitudes, respectively.

For each detected channel c, we select the three nearest eligible non-bad channels in the fixed three-dimensional electrode-coordinate system. Let $\mathcal{N}_c$ denote these donors and $d_{cj} = \|\mathbf{r}_c - \mathbf{r}_j\|_2$ their Euclidean distances. The repaired signal is

$$\widetilde{x}_c(t) = \sum_{j\in\mathcal{N}_c} w_{cj} x_j(t), \qquad w_{cj} = \frac{(d_{cj}+\epsilon)^{-1}}{\sum_{k\in\mathcal{N}_c}(d_{ck}+\epsilon)^{-1}}, \tag{14}$$

where $\epsilon > 0$ is a fixed numerical stabilizer and $|\mathcal{N}_c| = 3$. Donors are selected from the original non-bad channels, so repaired channels do not become donors for other repairs. Channels not marked bad are copied unchanged. If three valid donors are unavailable, the original channel is retained and the failed repair is flagged.

Following mdJPT (Zhang et al., 2025), we map the repaired signals to a fixed 60-channel montage with a common channel order. This step fills target electrodes absent from the native layout after repairing the observed channels. An available target channel is copied from the repaired native signal; a missing target is reconstructed by averaging its three predefined support channels. For target channel q,

$$y_{i,q}(t) = \begin{cases} \widetilde{x}_{i,\pi(q)}(t), & q \text{ is present}, \\ \dfrac{1}{3} \displaystyle\sum_{j \in \mathcal{P}_q} \widetilde{x}_{i,j}(t), & q \text{ is absent}, \end{cases} \tag{15}$$

where $\pi(q)$ identifies the corresponding native channel and $\mathcal{P}_q$ is the fixed three-channel support set. The support mapping is defined by the native electrode layout and the common target montage. Consequently, bad-channel repair uses normalized inverse-distance weights, whereas missing-target projection uses equal weights. The final segment is $Y_i \in \mathbb{R}^{60\times 2000}$, with separate masks identifying repaired native channels and projected target channels. Signals are exported with labels, sample and subject/session identifiers, valid lengths, quality flags, and preprocessing metadata.

B Implementation and Training Details

B.1 Training Hyperparameters and Loss Weights

PACE initializes each encoder from its official pretrained weights and jointly optimizes the encoder and shared projection head. Stage 2 inherits the source-validation-selected Stage 1 checkpoint, with the optimizer reset at stage entry. Table 6 summarizes the post-training settings; loss coefficients are specified separately in Table 8.

Table 6: PACE post-training hyperparameters. Learning rates apply jointly to the encoder and projection head. An anchor epoch measures valid training exposures, as defined below.

Setting	Value
Optimizer / weight decay	Adam / 10^{-4}
Learning rate, Stage 1 / Stage 2 (all backbones)	1.5×10^{-4} / 7.5×10^{-5}
Contrastive temperature	0.07
Distributed training	Four GPUs; averaged gradients; candidates remain local to each rank
Numerical precision	FP32 for all backbones
Maximum budget per stage	50 anchor epochs
Minimum budget before patience stopping	10 anchor epochs
Source-validation interval / patience	Approximately 2 anchor epochs / 5 non-improving validations
EA context size	At most 32 windows per subject block
Projection hidden / output width	256 / 128

An anchor epoch is the number of valid loss-anchor occurrences accumulated across ranks divided by the number of eligible source-training windows. EA context windows supply alignment statistics, while eligible anchors contribute to the contrastive objective. In Stage 2, occurrences from retained T/E objectives and both cross-dataset objectives all enter the numerator. The denominator is the union of eligible training windows across the active objectives, counting each window once; repeated loss-anchor occurrences count separately in the numerator.

The same checkpoint-selection rule is applied across all backbones using held-out source subjects.

LP standardizes each feature dimension using training-split statistics, with the standard deviation floored at 10^{-6}. FT uses a newly initialized classifier comprising LayerNorm, linear layers of widths 256 and 128, and a final target-class output layer. Each hidden layer uses ReLU followed by dropout with probability 0.2. Both modes optimize unweighted cross-entropy and stop after ten epochs without a relative training-loss improvement of at least 0.01. Checkpoints are selected by validation balanced accuracy, retaining the earlier epoch in a tie.

For every method and seed, we train with all three learning rates listed in Table 7, select the learning rate and checkpoint with the highest validation balanced accuracy, and evaluate the selected

Table 7: Downstream optimization settings for the foundation-model and PACE representations. The three learning-rate candidates apply to every method. Each target uses six repeated, subject-disjoint splits with seeds 42–47.

Setting	Linear probing (LP)	Full fine-tuning (FT)
Trainable components	Linear classifier	Representation and classifier
Optimizer	Adam	AdamW
Learning-rate candidates	2×10^{-4}, 10^{-3}, 5×10^{-3}	2×10^{-6}, 10^{-5}, 5×10^{-5}
Weight decay	10^{-4}	10^{-4}
Batch size	128	128
Maximum epochs	100	60
Gradient-norm clipping	—	1.0
Checkpoint selection	Validation balanced accuracy	Validation balanced accuracy

model once on the test split. The specific subject-level train/validation/test splits for each dataset are provided in the code.

Loss weights. Table 8 gives the loss coefficients for both training stages. Superscripts distinguish the weights of the within-dataset terms in the two stages.

Table 8: Loss weights for affective post-training. The coefficients of the same-paradigm and cross-paradigm terms multiply the average loss within each dataset-pair group.

Stage	Coefficient	Loss component	Value
1	$\lambda_T^{(1)}$	Within-dataset content/time correspondence	$1/2$
1	$\lambda_E^{(1)}$	Within-dataset emotion labels	$1/2$
2	$\lambda_T^{(2)}$	Retained content/time correspondence	$1/6$
2	$\lambda_E^{(2)}$	Retained native emotion labels	$1/6$
2	λ_s	Same-paradigm cross-dataset alignment	$1/3$
2	λ_c	Cross-paradigm cross-dataset alignment	$1/3$

In Stage 2, $1/|G_s|$ and $1/|G_c|$ average over the eligible dataset pairs in each group. Thus, λ_s and λ_c specify group weights; the coefficient of an individual pair loss is $\lambda_s/|G_s|$ or $\lambda_c/|G_c|$, respectively. This expression assumes both groups are nonempty and each included dataset pair has valid positive relations.

B.2 Representation, Alignment, and Transfer Protocol

CE, SupCon, VREx, and PACE use the same encoder–projector architecture and downstream protocol. Published model baselines retain their prescribed readouts, including their original projection heads where applicable. For an EEG window x_i from dataset d_i, the common representation and classification logits are

$$r_i = \mathcal{R}\Big(g_\phi\Big(\widetilde{H}_i\Big)\Big), \qquad \mathbf{a}_i = h_{d_i}(r_i), \tag{16}$$

where $\widetilde{H}_i$ denotes the encoder tokens after EA32, defined below, and $\mathcal{R}$ flattens the projected tokens and applies ℓ_2 normalization. All channel and temporal tokens are retained. The encoder–projector representation is frozen for LP, for which only a target linear classifier is fitted; FT jointly optimizes the transferred representation and target classifier.

B.2.1 Projection Head

The projection head applies the same two-layer MLP independently to every channel/time token. For an aligned token $\widetilde{\mathbf{h}} \in \mathbb{R}^D$,

$$g_\phi(\widetilde{\mathbf{h}}) = W_2\,\mathrm{GELU}(W_1\widetilde{\mathbf{h}} + b_1) + b_2, \qquad W_1 \in \mathbb{R}^{256\times D}, \quad W_2 \in \mathbb{R}^{128\times 256}. \tag{17}$$

The token width D is 200 for CBraMod, LaBraM, and CSBrain, and 512 for REVE. Thus, the respective projection dimensions are $200 \to 256 \to 128$ and $512 \to 256 \to 128$, with biases in

both linear layers. The projector has no dropout or additional normalization layers. Its parameters are shared across source datasets and objectives within each backbone, and Stage 2 continues the projector learned in Stage 1.

CBraMod, LaBraM, and CSBrain produce 60×10 tokens per input window; REVE produces 60×11 tokens from its internal patching of the same 10-s input. Flattening the projected tokens therefore yields 76,800 features for the first three backbones and 84,480 for REVE. The learned projector is retained during downstream transfer, frozen in LP and updated together with the encoder in FT.

B.2.2 Feature-Space EA32

EA32 performs regularized symmetric ZCA whitening in the encoder feature space before projection. Within each dataset, subject, and data partition, windows follow a fixed label-independent order and form blocks of at most 32 windows. The final shorter block is retained. All channel/time tokens in a block contribute observations for estimating its feature statistics.

Let $H_b \in \mathbb{R}^{M_b\times D}$ collect the encoder tokens of block b, where M_b counts tokens across its windows. We compute

$$\begin{aligned}\mu_b &= \frac{1}{M_b}\sum_{m=1}^{M_b} H_{b,m},\\ \Sigma_b &= \frac{(H_b-\mathbf{1}\mu_b)^\top(H_b-\mathbf{1}\mu_b)}{M_b-1},\\ s_b &= \max\!\left(\frac{\operatorname{tr}(\Sigma_b)}{D},10^{-8}\right), \qquad \Sigma_b^{\mathrm{reg}} = 0.9\Sigma_b + 0.1 s_b I.\end{aligned} \tag{18}$$

With eigendecomposition $\Sigma_b^{\mathrm{reg}} = Q_b\,\mathrm{diag}(\lambda_{b,j})Q_b^\top$, the aligned tokens are

$$\widetilde{H}_b = (H_b-\mathbf{1}\mu_b)Q_b\,\mathrm{diag}\Big(\max(\lambda_{b,j},10^{-5}s_b)^{-1/2}\Big)\,Q_b^\top. \tag{19}$$

The shrinkage coefficient is 0.1 and the relative eigenvalue floor is 10^{-5}. During optimization, means and whitening matrices are detached from the gradient graph; gradients propagate through the transformed encoder tokens and the projector.

At evaluation, EA32 estimates alignment statistics from blocks of up to 32 unlabeled windows from the same subject and partition. Blocks remain separate across training, validation, and test partitions, and never mix subjects or datasets. This blockwise statistical adaptation is distinct from LP feature standardization, whose statistics are fitted exclusively on the downstream training split.

B.2.3 Native and EA32 Baseline Configurations

Every baseline in the three comparison groups is evaluated in both its native configuration and a configuration augmented with EA32. These groups comprise EEG foundation models, emotion-pretrained models, and task-specific small models. The native configuration retains the model's original feature-processing pipeline; the EA32 configuration adds the feature-space alignment described above before downstream readout. For each baseline, target dataset, evaluation mode, and seed, the configuration, learning rate, and checkpoint are selected by validation balanced accuracy before the selected model is evaluated on the test split. This selection rule applies to all three baseline groups, while PACE uses EA32 throughout post-training and transfer. Foundation-model baselines with EA32 read out the aligned encoder features directly; PACE additionally retains its learned projection head.

B.3 Model Sources

The table below lists the original publications for the foundation backbones and comparison methods. All target performance values in our tables are results of the present evaluation.

B.4 Cross-Dataset Label Mapping and Pair Selection

This appendix specifies the predefined label mappings and pair-selection rules used for cross-dataset supervision in all Stage 2 variants. For bookkeeping, the notation distinguishes positive (P), negative

Table 9: Original publications for the foundation backbones and comparison methods.

Model	Original publication
LaBraM	Jiang et al. (2024)
CBraMod	Wang et al. (2025)
REVE	El Ouahidi et al. (2025)
CSBrain	Zhou et al. (2025)
EEGNet	Lawhern et al. (2018)
LGGNet	Ding et al. (2024)
DMMR	Wang et al. (2024)
CLISA	Shen et al. (2023)
FreqDGT	Li et al. (2025b)
mdJPT	Zhang et al. (2025)
MMM	Yi et al. (2023)
EMOD	Chen et al. (2026)
SupCon	Khosla et al. (2020)
VREx	Krueger et al. (2021)

(N), ambiguous (A), and unmapped (U) relations (PNAU). Only P/N pairs enter the cross-dataset objective; A/U pairs are excluded. Within-dataset supervision and target-dataset classification retain native labels throughout.

B.4.1 Shared Concepts and Exact Label Mapping

The shared vocabulary contains nine operational emotion concepts, indexed as follows: 0 neutral, 1 sadness, 2 fear, 3 disgust, 4 anger, 5 happiness, 6 joy, 7 inspiration, and 8 tenderness. Shared concepts define operational label correspondences across datasets.

Native labels are mapped using fixed exact aliases after lowercasing and trimming surrounding whitespace: `sad` maps to sadness and `happy` to happiness, while joy remains distinct from happiness. REFED uses four complete registered strings: `mvma (neutral)` maps to neutral, `lvla (sad)` to sadness, `lvha (fear)` to fear, and `hvha (happy)` to happiness. Its `hvla (relax)` label is unmapped; arbitrary parenthesized text is never used to infer a mapping.

Table 10 lists the concepts available to cross-dataset and cross-paradigm supervision. All other native categories remain U, including coarse valence labels, amusement, surprise, and calmness. The native within-dataset and downstream tasks are unchanged.

Table 10: Mapped concept IDs for fine-grained cross-dataset supervision. IDs follow the shared vocabulary above; a dash means no mapped categories.

Dataset	Mapped concept IDs
SEED	0
SEED-IV	0, 1, 2, 5
SEED-V	0, 1, 2, 3, 5
SEED-VII	0, 1, 2, 3, 4, 5
FACED	0, 1, 2, 3, 4, 6, 7, 8
REFED	0, 1, 2, 5
MC-vid, MC-ima	0, 1, 2, 3, 6, 7, 8
EAV	0, 1, 4, 5
ImaginedEmotion	1, 2, 3, 4, 5, 6
DEAP, MEEG	—

B.4.2 Categorical Eligibility Rules

Let $\ell_i = m_{d_i}(y_i)$ denote the mapped concept or $\perp$. We use a fixed semantic prior over emotion category definitions to determine cross-dataset pair eligibility. Negative eligibility follows the native

category distinctions of SEED-VII and FACED within the shared vocabulary:

$$\begin{aligned} \mathcal{V}_{\text{SEED-VII}} &= \{0,1,2,3,4,5\}, \\ \mathcal{V}_{\text{FACED}} &= \{0,1,2,3,4,6,7,8\}, \\ \mathcal{N} &= \bigcup_{d\in\{\text{SEED-VII,FACED}\}} \big\{\{a,b\} : a,b \in \mathcal{V}_d,\ a \neq b\big\}. \end{aligned} \tag{20}$$

The concept mapping and relation matrix remain fixed across all leave-one-dataset-out folds. They are specified from category definitions and are not estimated from target recordings or sample-level annotations. This corresponds to transfer to an unseen EEG dataset with a known downstream emotion taxonomy. Target recordings and sample-level annotations are held out throughout post-training, and all training pairs use source samples.

ImaginedEmotion does not contribute category distinctions to the matrix. Its six mapped concepts still participate in alignment when it is a source dataset. For samples from different datasets, the rule is

$$R_{ij} = \begin{cases} \text{U}, & \ell_i = \perp \text{ or } \ell_j = \perp, \\ \text{P}, & \ell_i, \ell_j \in \mathcal{V},\ \ell_i = \ell_j, \\ \text{N}, & \ell_i, \ell_j \in \mathcal{V},\ \{\ell_i, \ell_j\} \in \mathcal{N}, \\ \text{A}, & \text{otherwise.} \end{cases} \qquad d_i \neq d_j. \tag{21}$$

Writing $A_i = \{j : d_j \neq d_i,\ R_{ij} = \text{A}\}$ and $U_i = \{j : d_j \neq d_i,\ R_{ij} = \text{U}\}$, the active candidate set is

$$\mathcal{C}_i = P_i \cup N_i, \qquad (A_i \cup U_i) \cap \mathcal{C}_i = \varnothing. \tag{22}$$

Unmapped labels give U, even if their strings are identical, such as `positive` or `amusement`. Matching mapped concepts give P; distinct mapped concepts give N when supported by the registered category distinctions, and A otherwise. A therefore denotes an undetermined relation under this evidence scope. The matrix contains 33 unordered N pairs, three unordered A pairs, and nine diagonal P entries. U is absent because both axes contain only mapped concepts.

Table 11: Pair-selection matrix used for cross-dataset training. Column IDs correspond to the concept IDs in the first column. P: matching mapped concepts; N: eligible negatives; A: no registered category-distinction evidence, excluded from the contrastive denominator. Any pair involving an unmapped label is U and is excluded.

ID	Concept	0	1	2	3	4	5	6	7	8
0	neutral	P	N	N	N	N	N	N	N	N
1	sadness	N	P	N	N	N	N	N	N	N
2	fear	N	N	P	N	N	N	N	N	N
3	disgust	N	N	N	P	N	N	N	N	N
4	anger	N	N	N	N	P	N	N	N	N
5	happiness	N	N	N	N	N	P	A	A	A
6	joy	N	N	N	N	N	A	P	N	N
7	inspiration	N	N	N	N	N	A	N	P	N
8	tenderness	N	N	N	N	N	A	N	N	P

The three unordered A pairs are happiness–joy, happiness–inspiration, and happiness–tenderness. Distinctions among sadness, fear, disgust, and anger, and among joy, inspiration, and tenderness, are N under the registered evidence. Neutral forms N with every other mapped concept.

A legal dataset pair requires at least one shared mapped class. Queries enter the loss whenever they have a legal positive; Relation AUC additionally requires at least one legal negative. Nonshared mapped classes can remain negative candidates according to the matrix. The two dataset sides may have different sample counts. Each sampled side uses distinct episodes and prioritizes shared classes before covering other mapped classes. Target holdout, MC-family holdout, and subject eligibility restrictions continue to determine which source pairs are available in each run.

Only P/N relations contribute to the cross-dataset objective; A/U relations impose no similarity constraint. The same relation matrix applies to same-paradigm and cross-paradigm dataset pairs and to the representation diagnosis in Appendix E.1.

C Complete Transfer Results

Tables 12, 13, 14, and 15 report balanced accuracy and macro-F1 in percent. The benchmark baselines use subject-disjoint 25%/25%/50% train/validation/test splits with seeds 42–47. Windows from a subject remain in one partition. Inputs are 10-s, 60-channel EEG segments sampled at 200 Hz; standardization is fitted only on the current training subjects. LP freezes the backbone and trains a linear readout, whereas FT updates all model parameters. Each supplied benchmark cell reports the six-seed mean and population standard deviation ($\mathrm{ddof} = 0$).

For the benchmark baselines, Avg9 averages SEED, SEED-IV, SEED-V, SEED-VII, DEAP, REFED, FACED, EAV-L (listening), and MEEG equally within each seed, then reports the mean and population standard deviation across seeds. MC-V, MC-I, and Imag. are reference columns excluded from Avg9. The foundation-backbone and PACE panels retain Avg12, which weights all twelve targets equally; Avg9 and Avg12 summarize different target sets. In the benchmark panels, $*$ denotes a project-adapted supervised implementation, and † denotes MMM with the project adapter and input interface. FT macro-F1 is reported for foundation backbones and PACE only.

PACE denotes the complete sequential S1→S2 run. Stage ablations, AUROC results, and comparisons with alternative post-training objectives are provided in Appendix D.

C.1 Full Fine-Tuning

Table 12: Balanced accuracy with FT (%). Benchmark baselines use Avg9; foundation backbones and PACE use Avg12. Bold baseline entries indicate the best benchmark baseline in each column.

Benchmark baselines: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
EEGNet	51.57±5.34	29.34±3.43	25.19±2.12	**20.69±0.98**	37.36±0.70	**26.30±1.59**
TSception	47.52±3.20	29.36±2.28	**25.88±2.24**	18.84±0.47	19.39±0.45	19.25±1.38
LGGNet	51.87±2.66	32.22±1.61	25.60±1.30	19.90±0.31	30.58±1.07	21.20±1.64
DMMR*	47.04±2.16	32.06±1.16	25.25±1.73	19.11±0.84	12.62±1.02	17.85±0.91
HEDN*	45.27±1.80	32.55±2.15	23.50±0.91	18.87±0.87	11.09±0.05	16.38±0.83
CLISA*	49.78±2.98	29.28±2.92	23.40±0.17	18.68±0.50	**40.79±1.43**	15.74±0.79
FreqDGT*	46.08±2.20	30.64±1.11	23.22±0.84	17.65±0.79	14.48±0.86	15.60±0.87
MMM†	49.58±3.78	30.12±2.79	24.68±3.06	18.05±2.07	15.38±0.39	15.38±0.82
mdJPT	**52.55±3.66**	**33.59±1.86**	25.76±2.10	20.51±0.44	19.56±1.30	19.57±1.59
EMOD	45.49±0.99	28.69±2.70	23.33±1.95	17.16±1.42	24.12±6.36	17.76±1.36

Benchmark baselines: REFED through Imag.; Avg9

Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg9
EEGNet	22.20±0.54	**52.88±0.32**	32.95±1.27	61.53±2.52	14.41±0.69	8.67±1.37	**37.08±0.86**
TSception	25.51±1.67	51.08±1.15	**33.83±1.43**	**64.13±4.54**	14.33±0.69	8.55±0.90	35.06±0.33
LGGNet	24.82±0.46	51.92±1.42	31.09±1.49	60.41±1.82	14.13±0.62	7.81±0.42	36.49±0.51
DMMR*	**25.79±1.82**	50.39±1.35	32.18±1.12	62.30±3.85	14.84±1.26	8.23±1.11	34.08±0.26
HEDN*	20.79±1.13	49.83±0.53	30.08±2.91	62.01±5.99	14.11±0.63	7.50±1.22	32.67±0.46
CLISA*	21.86±1.15	52.67±1.00	25.51±2.68	52.85±1.29	14.56±0.78	6.96±0.27	34.98±0.63
FreqDGT*	23.82±0.62	50.08±1.35	28.77±1.03	58.34±2.78	**14.90±0.84**	7.23±0.74	32.56±0.22
MMM†	22.16±2.11	50.76±2.46	33.17±1.83	57.30±7.70	14.76±1.40	6.84±1.11	33.46±1.23
mdJPT	23.72±1.35	51.61±0.72	31.36±1.42	63.05±3.90	13.73±1.03	7.36±0.60	35.74±0.33
EMOD	23.84±1.16	50.42±0.85	30.70±1.47	61.60±3.61	14.42±1.01	**9.19±0.47**	33.93±1.26

EEG foundation models: SEED through MC-V						
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod	53.35±3.42	36.38±1.76	29.68±1.20	23.99±0.87	33.45±0.66	20.32±2.03
+ PACE	**64.09**±1.80	**44.04**±1.74	**38.56**±1.50	**26.50**±1.01	**54.40**±0.95	**34.19**±1.87
REVE	61.25±2.04	41.28±2.85	33.49±2.94	24.22±1.10	44.05±2.28	29.94±1.10
+ PACE	**69.30**±2.16	**46.03**±2.10	**42.81**±1.36	**29.54**±0.82	37.07±1.20	**34.88**±1.72
CSBrain	53.29±2.15	35.13±1.85	29.73±1.06	22.56±1.16	33.03±0.60	19.29±0.99
+ PACE	**58.14**±3.19	**39.67**±1.52	**34.53**±1.91	**25.26**±1.05	**44.69**±0.68	**28.35**±1.13
LaBraM	60.08±3.68	39.65±1.93	32.77±2.11	23.34±1.00	37.54±0.94	18.93±1.24
+ PACE	**65.21**±1.85	**41.88**±2.06	**37.80**±1.05	**27.58**±1.75	**38.60**±2.00	**26.59**±1.45

EEG foundation models: REFED through Imag.; Avg12							
Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg12
CBraMod	26.17±0.33	52.45±0.61	36.21±1.33	61.01±2.35	13.98±0.79	8.71±0.55	32.98
+ PACE	**31.54**±1.82	**58.10**±0.58	**39.68**±1.03	**64.35**±1.51	**15.39**±0.72	**9.53**±1.75	**40.03**
REVE	32.60±2.18	55.34±0.67	43.15±0.80	67.64±3.82	15.53±1.11	10.94±1.49	38.29
+ PACE	**37.65**±2.10	**56.55**±0.96	**44.02**±1.46	**68.23**±3.78	14.68±0.19	10.53±1.12	**40.94**
CSBrain	27.26±0.96	52.22±1.29	35.22±1.66	61.76±1.76	14.63±0.92	8.34±0.78	32.70
+ PACE	**28.20**±1.47	**55.85**±1.15	**36.62**±1.21	**64.78**±1.90	**14.81**±1.04	**8.92**±0.82	**36.65**
LaBraM	29.11±1.94	54.06±1.11	38.20±0.89	65.73±2.43	15.67±0.79	9.67±0.80	35.40
+ PACE	**32.81**±1.41	**56.73**±1.56	**41.31**±1.19	**67.21**±4.19	14.47±0.79	9.19±1.04	**38.28**

Table 13: Macro-F1 with FT (%) for EEG foundation backbones and PACE. Avg12 averages all twelve targets.

EEG foundation models: SEED through MC-V						
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod	53.17±3.46	36.28±1.72	29.22±1.53	23.49±0.90	33.59±0.65	20.02±2.09
+ PACE	**63.91**±1.74	**43.73**±1.69	**37.37**±1.44	**25.81**±1.19	**54.00**±0.98	**33.20**±1.75
REVE	60.91±1.71	40.76±2.94	32.56±3.38	23.92±1.15	43.98±2.10	29.17±1.26
+ PACE	**69.30**±2.09	**45.78**±2.01	**42.08**±1.32	**29.09**±0.89	36.50±1.12	**33.62**±1.31
CSBrain	52.89±2.04	34.81±1.81	28.96±1.14	22.05±0.97	32.96±0.56	19.14±1.02
+ PACE	**57.73**±3.16	**38.69**±2.13	**33.40**±2.01	**24.19**±1.41	**44.49**±0.68	**27.57**±0.88
LaBraM	60.00±3.56	39.09±2.48	32.68±2.06	23.13±0.98	37.45±0.92	18.60±1.09
+ PACE	**65.12**±1.69	**41.67**±2.23	**37.22**±1.20	**26.92**±2.15	**38.18**±2.03	**25.96**±1.24

EEG foundation models: REFED through Imag.; Avg12							
Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg12
CBraMod	23.96±1.38	51.15±1.92	35.79±1.28	60.96±2.35	11.59±1.71	6.81±1.17	32.17
+ PACE	**30.25**±2.64	**57.93**±0.63	**38.06**±0.81	**63.57**±2.29	**13.54**±2.02	**7.61**±1.99	**39.08**
REVE	31.80±1.83	54.25±1.18	42.03±1.55	67.47±3.74	14.16±1.67	8.34±2.38	37.45
+ PACE	**37.29**±1.98	**56.32**±0.95	**43.39**±1.66	**68.19**±3.85	**14.40**±0.17	**9.34**±1.47	**40.44**
CSBrain	26.57±1.16	51.45±2.18	34.21±1.37	61.67±1.79	13.70±1.36	7.63±0.59	32.17
+ PACE	**27.69**±1.39	**54.85**±2.28	**35.87**±1.35	**64.74**±1.86	**13.71**±1.16	**7.75**±1.36	**35.89**
LaBraM	28.02±2.17	53.65±1.47	37.67±0.90	65.62±2.44	15.16±0.99	8.30±1.15	34.95
+ PACE	**32.43**±1.09	**56.36**±2.02	**40.52**±1.63	**67.19**±4.22	14.25±0.75	7.03±0.41	**37.74**

C.2 Linear Probing

Table 14: Balanced accuracy with LP (%). Benchmark baselines use Avg9; foundation backbones and PACE use Avg12. Bold baseline entries indicate the best benchmark baseline in each column.

Benchmark baselines: SEED through MC-V						
Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
MMM†	44.24±2.30	29.44±1.89	24.71±2.30	17.22±0.63	15.26±0.46	16.25±1.06
mdJPT	**49.05±3.33**	**33.02±0.78**	**27.03±1.29**	**20.22±0.29**	**22.08±0.54**	**17.95±0.90**
EMOD	39.50±0.50	30.77±1.38	23.32±1.09	16.97±0.83	20.26±0.45	15.41±0.81

Benchmark baselines: REFED through Imag.; Avg9

Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg9
MMM†	**26.18±1.28**	49.97±0.57	29.98±1.04	**64.11±3.78**	**14.56±0.80**	7.48±0.98	33.46±0.25
mdJPT	24.41±1.05	**52.20±0.93**	**30.56±1.38**	62.31±3.15	14.42±0.93	7.97±0.42	**35.65±0.52**
EMOD	23.09±0.84	51.18±1.88	28.07±1.15	54.12±2.30	14.21±1.02	**8.26±0.46**	31.92±0.55

EEG foundation models: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod	53.78±1.80	35.70±1.08	29.65±0.97	23.31±0.83	35.39±0.81	22.55±1.31
+ PACE	**60.41**±1.48	**40.41**±0.86	**33.31**±1.00	**23.48**±0.87	**46.43**±1.23	**30.09**±1.91
REVE	60.56±3.18	37.39±1.92	31.26±1.65	23.83±1.54	35.71±1.46	28.42±1.54
+ PACE	**67.56**±2.00	**44.12**±0.97	**41.37**±1.17	**28.96**±0.64	**43.56**±0.68	**39.24**±1.64
CSBrain	46.72±2.06	30.86±0.89	24.28±0.60	19.21±0.36	23.81±0.77	16.73±1.58
+ PACE	**54.23**±1.49	**35.97**±0.66	**31.87**±0.96	**22.81**±0.72	**45.14**±0.83	**27.45**±1.09
LaBraM	56.48±3.29	37.01±0.97	30.63±1.79	23.48±0.67	34.74±0.35	19.40±1.49
+ PACE	**64.29**±2.29	**40.00**±1.86	**37.31**±1.59	**26.32**±0.78	**45.01**±0.82	**28.58**±1.71

EEG foundation models: REFED through Imag.; Avg12

Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg12
CBraMod	26.15±0.95	53.19±0.97	34.46±0.64	62.51±1.75	14.28±0.71	9.09±0.50	33.34
+ PACE	**29.67**±1.18	**55.42**±0.45	**36.33**±1.08	**62.63**±2.02	**15.41**±1.35	8.76±1.20	**36.86**
REVE	31.17±1.38	53.55±1.32	39.23±0.91	66.63±3.94	15.57±1.59	9.61±1.17	36.08
+ PACE	**34.22**±0.92	**56.32**±0.91	**39.87**±0.50	**68.16**±3.83	15.23±1.04	9.59±0.65	**40.68**
CSBrain	24.04±0.88	51.61±0.88	29.35±0.78	56.61±1.64	14.84±1.61	7.51±0.45	28.80
+ PACE	**26.58**±0.96	**53.84**±2.08	**32.82**±1.08	**61.65**±2.65	14.39±0.45	**8.02**±0.90	**34.56**
LaBraM	26.78±1.16	53.49±0.88	35.17±0.45	65.19±2.40	14.25±0.97	9.01±0.73	33.80
+ PACE	**32.00**±1.72	**55.58**±0.73	**37.30**±0.89	65.07±3.10	**14.83**±0.75	**9.36**±0.93	**37.97**

Table 15: Macro-F1 with LP (%). Benchmark baselines use Avg9; foundation backbones and PACE use Avg12. Bold baseline entries indicate the best benchmark baseline in each column.

Benchmark baselines: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
MMM†	41.48±4.36	25.98±4.75	21.79±2.28	13.06±0.92	13.87±0.88	15.45±1.25
mdJPT	**47.99±3.66**	**32.51±0.86**	**26.24±1.53**	**19.40±0.25**	**21.76±0.58**	**17.68±0.86**
EMOD	36.24±2.24	29.66±1.46	21.24±1.83	12.07±1.37	19.93±0.45	14.27±1.86

Benchmark baselines: REFED through Imag.; Avg9

Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg9
MMM†	**24.46±1.71**	48.79±1.37	25.31±2.29	**63.84±3.98**	12.57±1.20	6.11±1.73	30.95±0.82
mdJPT	22.37±1.58	**51.00±2.23**	**29.73±1.02**	62.08±3.23	**14.23±0.98**	6.87±0.60	**34.79±0.59**
EMOD	14.97±1.70	49.85±3.15	24.76±2.43	51.26±6.62	12.79±1.43	**7.45±0.58**	28.88±1.47

EEG foundation models: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod	53.56±1.59	35.21±1.35	28.98±1.01	23.00±1.11	35.23±0.73	22.38±1.21
+ PACE	**60.36**±1.28	**39.76**±0.97	**32.60**±0.97	**23.03**±0.95	**46.22**±1.30	**29.65**±1.85
REVE	60.29±3.03	36.37±2.64	29.94±1.78	23.35±1.85	35.42±1.55	27.66±1.50
+ PACE	**67.61**±1.73	**43.21**±1.24	**40.43**±1.09	**28.54**±0.81	**43.35**±0.61	**37.97**±1.80
CSBrain	46.77±2.07	30.63±0.85	23.73±0.43	19.08±0.40	23.72±0.83	16.64±1.58
+ PACE	**54.09**±1.40	**35.30**±1.16	**30.73**±0.87	**22.34**±0.72	**44.96**±0.89	**26.88**±1.08
LaBraM	56.16±3.13	36.56±1.03	29.89±1.79	23.19±0.75	34.57±0.37	19.29±1.42
+ PACE	**64.26**±2.15	**39.10**±1.53	**36.18**±1.37	**25.85**±0.99	**44.78**±0.89	**28.05**±1.71

EEG foundation models: REFED through Imag.; Avg12							
Model	REFED	DEAP	EAV-L	MEEG	MC-I	Imag.	Avg12
CBraMod	26.07±1.05	52.79±1.11	33.95±0.68	62.48±1.74	14.08±0.70	8.76±0.57	33.04
+ PACE	**29.47**±1.23	**54.84**±0.55	**34.78**±0.54	62.03±2.45	**14.88**±0.99	7.91±0.95	**36.29**
REVE	30.81±1.67	52.51±2.43	38.12±0.97	66.42±4.11	15.13±1.67	9.00±1.04	35.42
+ PACE	**33.80**±0.88	**55.91**±1.03	**39.19**±0.67	**68.10**±3.88	14.88±0.99	**9.10**±0.70	**40.17**
CSBrain	23.96±0.97	51.03±1.11	29.10±0.75	56.57±1.64	14.73±1.66	7.36±0.40	28.61
+ PACE	**26.28**±1.01	**53.30**±2.00	**32.14**±1.17	**61.57**±2.65	14.13±0.48	**7.71**±0.82	**34.12**
LaBraM	26.50±1.26	52.88±1.31	34.53±0.65	65.10±2.43	14.08±0.86	8.64±0.84	33.45
+ PACE	**31.49**±1.77	**55.20**±1.02	**36.53**±1.03	64.99±3.16	**14.57**±0.84	**8.90**±1.11	**37.49**

D Extended Ablation Studies

This appendix collects the progressive-stage ablation, its complementary AUROC results, and comparisons with alternative post-training objectives. All evaluations use the same twelve held-out targets and six subject splits.

D.1 Progressive Learning Stages

Table 16 reports the CBraMod stage ablation on the twelve original-protocol targets. The backbone is the pretrained R0+EA32 reference; S1-only and S2-only are single-stage variants, each trained for 100 epochs, and PACE is the complete sequential S1→S2 run. Entries report test balanced accuracy averaged over six seeds (42–47), with population standard deviations where available. The paired rows report PACE minus the indicated reference in percentage points. Avg12 differences use the displayed aggregate means, matching the main-text convention. Under FT, PACE improves on S1-only on 11 of the 12 targets, with a mean gain of 1.93 percentage points.

Table 16: Complete CBraMod stage ablation. Balanced accuracy (%). †: means follow the main-text table; corresponding split standard deviations were not supplied and are omitted.

FT: SEED through MC-V						
Variant	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod R0 + EA32	53.35±3.42	36.38±1.76	29.68±1.20	23.99±0.87	33.45±0.66	20.32±2.03
S1-only (T+E)	58.98±1.04	40.38±2.04	35.21±1.33	25.89±1.11	49.23±0.77	32.59±1.62
S2-only (no S1)	50.60±1.33	33.02±1.76	28.43±1.59	21.26±1.02	41.78±1.82	19.45±3.75
S1→S2 (PACE)	**64.09±1.80**	**44.04±1.74**	**38.56±1.50**	**26.50±1.01**	**54.40±0.95**	**34.19±1.87**
Δ PACE–S1 (pp)	**+5.11**	**+3.66**	**+3.35**	**+0.61**	**+5.17**	**+1.60**
Δ PACE–R0 (pp)	**+10.74**	**+7.66**	**+8.88**	**+2.51**	**+20.95**	**+13.87**
Δ PACE–S2 (pp)	**+13.49**	**+11.02**	**+10.13**	**+5.24**	**+12.62**	**+14.74**

FT: REFED through Imag.							
Variant	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
CBraMod R0 + EA32	26.17±0.33	52.45±0.61	36.21±1.33	61.01±2.35	13.98±0.79	8.71±0.55	32.98±0.63
S1-only (T+E)	30.58†	56.00†	39.38±1.47	64.15±2.66	15.25±0.39	**9.59±0.85**	38.10
S2-only (no S1)	24.43±3.12	52.14±1.56	34.75±1.80	59.64±3.78	14.00±0.78	9.25±1.45	32.40±0.79
S1→S2 (PACE)	**31.54±1.82**	**58.10±0.58**	**39.68±1.03**	**64.35±1.51**	**15.39±0.72**	9.53±1.75	**40.03**±0.42
Δ PACE–S1 (pp)	**+0.96**	**+2.10**	**+0.30**	**+0.20**	**+0.14**	−0.06	**+1.93**
Δ PACE–R0 (pp)	**+5.37**	**+5.65**	**+3.47**	**+3.34**	**+1.41**	**+0.82**	**+7.05**
Δ PACE–S2 (pp)	**+7.11**	**+5.96**	**+4.93**	**+4.71**	**+1.39**	**+0.28**	**+7.63**

LP: SEED through MC-V						
Variant	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
CBraMod R0 + EA32	53.78±1.80	35.70±1.08	29.65±0.97	23.31±0.83	35.39±0.81	22.55±1.31
S1-only (T+E)	56.17±2.31	37.13±1.37	32.95±1.66	**23.98±0.47**	43.32±1.12	**32.18±3.00**
S2-only (no S1)	51.27±2.20	32.45±1.16	28.19±0.76	21.05±0.58	29.02±0.29	20.10±1.46
S1→S2 (PACE)	**60.41±1.48**	**40.41±0.86**	**33.31±1.00**	23.48±0.87	**46.43±1.23**	30.09±1.91
Δ PACE–S1 (pp)	**+4.24**	**+3.28**	**+0.36**	−0.50	**+3.11**	−2.09
Δ PACE–R0 (pp)	**+6.63**	**+4.71**	**+3.66**	**+0.17**	**+11.04**	**+7.54**
Δ PACE–S2 (pp)	**+9.14**	**+7.96**	**+5.12**	**+2.43**	**+17.41**	**+9.99**

LP: REFED through Imag. Variant	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
CBraMod R0 + EA32	26.15±0.95	53.19±0.97	34.46±0.64	62.51±1.75	14.28±0.71	**9.09±0.50**	33.34±0.19
S1-only (T+E)	**29.75±1.78**	**55.43±0.49**	35.72±0.60	61.53±2.55	15.10±0.77	8.62±0.57	35.99±0.37
S2-only (no S1)	24.86±0.91	52.12±0.62	34.39±0.76	59.87±1.77	14.88±0.96	8.94±0.83	31.43±0.40
S1→S2 (PACE)	29.67±1.18	55.42±0.45	**36.33±1.08**	**62.63±2.02**	**15.41±1.35**	8.76±1.20	**36.86**±0.43
Δ PACE–S1 (pp)	−0.08	−0.01	**+0.61**	**+1.10**	**+0.31**	**+0.14**	**+0.87**
Δ PACE–R0 (pp)	**+3.52**	**+2.23**	**+1.87**	**+0.12**	**+1.13**	−0.33	**+3.52**
Δ PACE–S2 (pp)	**+4.81**	**+3.30**	**+1.94**	**+2.76**	**+0.53**	−0.18	**+5.43**

D.2 CBraMod AUROC

Table 17 complements the transfer results in Appendix C and the stage ablation above with AUROC for the same CBraMod runs. These values use the same six subject splits (seeds 42–47), population standard deviation, and twelve original-protocol targets. PACE denotes the complete sequential S1→S2 run.

Table 17: CBraMod AUROC for FT and LP (%).

FT: SEED through MC-V Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	72.41±2.95	62.97±1.06	61.91±1.44	63.56±1.11	75.34±0.55	58.00±1.98
S1	81.38±1.36	69.75±0.84	70.90±1.21	66.91±0.86	87.73±0.45	72.14±1.53
PACE	**82.23±1.74**	**69.70±1.19**	**70.33±1.38**	**66.77±0.99**	**87.67±0.37**	**71.35±1.67**

FT: REFED through Imag. Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	57.91±0.77	53.44±0.66	67.89±1.30	65.33±3.27	49.97±0.71	55.52±2.02	62.02
S1	64.39±1.17	60.99±0.81	72.06±0.73	68.48±3.71	51.30±0.67	57.30±2.92	68.61
PACE	**64.15±1.23**	**61.28±0.26**	**71.77±0.93**	**69.11±2.72**	**51.24±0.98**	**56.82±2.26**	**68.53**

LP: SEED through MC-V Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	72.50±1.75	61.74±0.85	61.00±1.14	61.31±0.74	75.98±0.55	60.23±1.61
S1	77.34±2.38	65.54±1.11	65.53±1.57	61.33±1.53	83.20±0.59	68.73±2.18
PACE	**78.67±1.52**	**65.84±0.83**	**64.54±0.96**	**61.95±1.02**	**83.17±0.37**	**68.01±1.55**

LP: REFED through Imag. Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	57.39±0.89	54.64±0.93	66.07±0.78	66.78±2.33	50.35±0.46	54.47±1.29	61.87
S1	60.77±1.52	57.36±0.44	66.22±1.29	64.62±3.31	51.42±1.03	54.51±1.97	64.71
PACE	**61.09±1.36**	**57.50±0.72**	**67.05±0.85**	**65.43±3.34**	**51.20±1.08**	**53.08±0.92**	**64.79**

D.3 Comparison with Alternative Post-Training Methods

The CBraMod comparison uses the same twelve held-out targets and six-seed readout protocol. CE denotes supervised cross-entropy post-training, SupCon supervised contrastive learning, and VREx variance risk extrapolation. PACE is the complete S1→S2 run. The displayed Δ row is PACE minus the strongest CE/SupCon/VREx baseline for each target.

Table 18: Complete CBraMod post-training comparison. Balanced accuracy (%).

FT: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	53.35±3.42	36.38±1.76	29.68±1.20	23.99±0.87	33.45±0.66	20.32±2.03
CE	58.11±1.56	37.51±2.26	31.73±2.18	23.50±0.66	33.99±1.03	20.59±1.23
SupCon	56.93±1.99	37.61±1.36	31.83±1.38	23.95±0.92	36.88±0.61	25.41±1.65
VREx	55.94±2.09	36.28±1.73	31.27±1.26	23.42±1.45	38.76±0.88	25.63±1.27
PACE	64.09±1.80	44.04±1.74	38.56±1.50	26.50±1.01	54.40±0.95	34.19±1.87
Δ vs. best baseline (pp)	**+5.98**	**+6.43**	**+6.73**	**+2.55**	**+15.64**	**+8.56**

FT: REFED through Imag.

Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	26.17±0.33	52.45±0.61	36.21±1.33	61.01±2.35	13.98±0.79	8.71±0.55	32.98±0.63
CE	27.91±1.79	53.18±1.04	38.74±1.19	63.03±3.83	14.86±0.72	9.22±0.73	34.36±0.18
SupCon	27.02±1.65	53.21±0.59	36.77±1.39	63.26±3.07	15.04±0.73	9.09±0.96	34.75±0.56
VREx	26.70±1.56	53.12±0.93	37.47±1.33	62.40±4.88	14.80±0.56	8.63±1.27	34.54±0.22
PACE	31.54±1.82	58.10±0.58	39.68±1.03	64.35±1.51	15.39±0.72	9.53±1.75	**40.03±0.42**
Δ vs. best baseline (pp)	**+3.63**	**+4.89**	**+0.94**	**+1.09**	**+0.35**	**+0.31**	**+5.28**

LP: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	53.78±1.80	35.70±1.08	29.65±0.97	23.31±0.83	35.39±0.81	22.55±1.31
CE	53.49±1.79	34.55±1.84	29.59±1.31	20.51±1.37	26.14±0.83	18.79±0.54
SupCon	55.01±2.39	34.71±0.88	29.07±1.68	21.68±0.83	34.44±0.63	24.60±2.03
VREx	53.27±1.80	33.18±0.86	29.63±1.38	21.30±1.15	27.51±0.42	23.08±1.06
PACE	60.41±1.48	40.41±0.86	33.31±1.00	23.48±0.87	46.43±1.23	30.09±1.91
Δ vs. best baseline (pp)	**+6.63**	**+5.70**	**+3.68**	**+1.80**	**+11.99**	**+5.49**

LP: REFED through Imag.

Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	26.15±0.95	53.19±0.97	34.46±0.64	62.51±1.75	14.28±0.71	9.09±0.50	33.34±0.19
CE	27.21±1.82	52.41±1.13	35.86±0.78	64.04±4.12	14.93±1.21	8.08±0.65	32.13±0.31
SupCon	25.46±1.13	53.69±0.50	34.34±0.93	61.25±2.77	14.66±0.43	9.32±0.65	33.19±0.38
VREx	26.02±1.77	53.02±0.83	34.73±0.90	63.27±3.47	14.41±0.82	7.93±0.67	32.28±0.16
PACE	29.67±1.18	55.42±0.45	36.33±1.08	62.63±2.02	15.41±1.35	8.76±1.20	**36.86±0.43**
Δ vs. best baseline (pp)	**+2.46**	**+1.73**	**+0.47**	−1.41	**+0.48**	−0.56	**+3.67**

Table 19: Complete CBraMod post-training comparison. Macro-F1 (%).

FT: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	53.17±3.46	36.28±1.72	29.22±1.53	23.49±0.90	33.59±0.65	20.02±2.09
CE	57.94±1.31	37.10±2.49	30.30±2.72	23.25±0.62	33.70±1.08	20.41±1.17
SupCon	56.54±1.70	37.51±1.12	31.29±1.34	23.52±0.92	36.99±0.60	24.98±1.45
VREx	55.69±2.10	35.94±2.18	30.67±1.19	23.17±1.40	38.78±0.90	25.39±1.24
PACE	63.91±1.74	43.73±1.69	37.37±1.44	25.81±1.19	54.00±0.98	33.20±1.75
Δ vs. best baseline (pp)	**+5.97**	**+6.22**	**+6.08**	**+2.29**	**+15.22**	**+7.81**

FT: REFED through Imag.

Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	23.96±1.38	51.15±1.92	35.79±1.28	60.96±2.35	11.59±1.71	6.81±1.17	32.17
CE	26.31±2.63	52.51±1.56	38.22±1.08	62.49±4.47	14.00±1.48	7.58±1.50	33.65
SupCon	25.40±2.39	52.77±0.79	35.31±1.81	63.02±3.44	13.06±1.59	7.52±1.52	33.99
VREx	25.63±2.00	52.79±1.03	36.76±1.32	61.07±7.71	13.18±2.12	7.32±1.05	33.87
PACE	30.25±2.64	57.93±0.63	38.06±0.81	63.57±2.29	13.54±2.02	7.61±1.99	**39.08**
Δ vs. best baseline (pp)	**+3.94**	**+5.14**	−0.16	**+0.55**	−0.46	**+0.03**	**+5.09**

LP: SEED through MC-V

Model	SEED	SEED-IV	SEED-V	SEED-VII	FACED	MC-V
Backbone	53.56±1.59	35.21±1.35	28.98±1.01	23.00±1.11	35.23±0.73	22.38±1.21
CE	53.36±1.82	34.05±1.73	28.93±1.69	19.94±1.44	25.99±0.84	18.48±0.44
SupCon	54.81±2.19	34.20±1.07	28.25±1.59	21.42±0.88	34.33±0.63	24.20±1.93
VREx	53.12±1.67	32.72±1.00	28.89±1.49	21.02±1.08	27.37±0.39	22.90±1.12
PACE	60.36±1.28	39.76±0.97	32.60±0.97	23.03±0.95	46.22±1.30	29.65±1.85
Δ vs. best baseline (pp)	**+5.55**	**+5.56**	**+3.62**	**+1.61**	**+11.89**	**+5.45**

LP: REFED through Imag.

Model	REFED	DEAP	EAV	MEEG	MC-I	Imag.	Avg12
Backbone	26.07±1.05	52.79±1.11	33.95±0.68	62.48±1.74	14.08±0.70	8.76±0.57	33.04
CE	27.01±1.72	52.01±1.43	35.25±0.99	63.82±4.31	14.65±1.23	7.61±0.57	31.76
SupCon	25.43±1.10	53.09±0.84	33.69±0.98	61.20±2.78	14.46±0.46	8.84±0.68	32.83
VREx	25.87±1.85	52.30±1.03	34.17±1.01	63.17±3.53	14.17±0.92	7.63±0.68	31.94
PACE	29.47±1.23	54.84±0.55	34.78±0.54	62.03±2.45	14.88±0.99	7.91±0.95	**36.29**
Δ vs. best baseline (pp)	**+2.46**	**+1.75**	−0.47	−1.79	**+0.23**	**+0.30**	**+3.47**

E Cross-Context Representation Diagnosis

We compare pretrained backbones with their PACE counterparts to assess how frozen EEG representations preserve emotion relations across subjects, datasets, and elicitation paradigms. Each target dataset is held out from its checkpoint's post-training sources; no downstream classifier is fitted.

E.1 Evaluation protocol

Evaluation contexts. We evaluate four contexts: within-subject (WS), which compares different trials from the same subject and dataset; cross-subject (CS), which compares different subjects within the same dataset; cross-dataset (CD), which compares datasets sharing the same elicitation paradigm; and cross-paradigm (CP), which compares datasets using different elicitation paradigms.

Sampling and scoring. All systems share queries, galleries, and three segment-sampling seeds. We sample at most four query trials per eligible class and subject, with up to four validation subjects per gallery dataset for CS/CD/CP; WS uses other trials from the query subject. Diagnosis uses the same label mapping and P/N/A/U relation matrix as training (Appendix B.4). Matching mapped emotions form positives (P), eligible distinct mapped emotions form negatives (N), and ambiguous (A) or unmapped (U) relations are excluded. The all-valid evaluation excludes the query trial and identifiable shared stimuli or events. Relation AUC measures how often cosine similarity ranks a positive above a negative, with half credit for ties. We report $100 \times \mathrm{AUC}$, with chance at 50.

Aggregation and uncertainty. We average seeds, gallery subjects, gallery datasets, and negative classes, then query trials within class, classes within subject, subjects within target, and finally targets, using equal weights at each level. WS/CS/CD/CP include 7/8/8/9 eligible targets. PACE gains use matched scoring units. Pointwise 95% intervals come from 1,000 query-subject cluster-bootstrap replicates with fixed galleries and equal target weights; they quantify sampling uncertainty and are unadjusted for multiple comparisons.

E.2 MACRO-AVERAGED RESULTS

Tables 20 and 21 report AUC and paired PACE gains. PACE improves all four contexts across all four backbones, with every paired 95% interval above zero. Gains reach 5.51 AUC points for cross-dataset transfer and 3.72 points for cross-paradigm transfer.

Table 20: All-valid diagnosis: AUC with 95% intervals. WS: Within-subject; CS: Cross-subject; CD: Cross-dataset; CP: Cross-paradigm.

Backbone	Readout	WS	CS	CD	CP
CBraMod	Backbone	57.75 [56.84,58.69]	51.95 [51.56,52.34]	50.59 [50.32,50.89]	49.94 [49.66,50.21]
CBraMod	PACE	61.08 [59.81,62.40]	55.18 [54.66,55.74]	53.68 [53.14,54.19]	52.68 [52.23,53.16]
LaBraM	Backbone	57.71 [56.75,58.65]	52.26 [51.89,52.66]	50.08 [49.74,50.39]	49.75 [49.46,50.04]
LaBraM	PACE	60.66 [59.38,61.84]	55.19 [54.61,55.75]	54.54 [53.98,55.19]	53.43 [52.97,53.90]
REVE	Backbone	58.51 [57.59,59.38]	52.16 [51.86,52.46]	50.32 [50.03,50.64]	50.74 [50.48,51.00]
REVE	PACE	61.89 [60.74,63.05]	55.04 [54.55,55.53]	55.83 [55.14,56.48]	54.46 [53.81,55.11]
CSBrain	Backbone	54.95 [54.05,55.87]	51.02 [50.65,51.40]	50.29 [49.96,50.60]	49.05 [48.79,49.30]
CSBrain	PACE	58.93 [57.80,60.00]	53.01 [52.55,53.48]	52.54 [52.00,53.07]	52.21 [51.69,52.81]

Table 21: All-valid paired PACE gains over the backbone with 95% intervals. WS: Within-subject; CS: Cross-subject; CD: Cross-dataset; CP: Cross-paradigm.

Backbone	ΔWS	ΔCS	ΔCD	ΔCP
CBraMod	+3.34 [+2.08,+4.38]	+3.24 [+2.69,+3.76]	+3.09 [+2.51,+3.64]	+2.74 [+2.29,+3.19]
LaBraM	+2.95 [+2.12,+3.82]	+2.92 [+2.33,+3.46]	+4.46 [+3.83,+5.22]	+3.68 [+3.25,+4.14]
REVE	+3.38 [+2.50,+4.24]	+2.88 [+2.41,+3.38]	+5.51 [+4.78,+6.17]	+3.72 [+3.08,+4.34]
CSBrain	+3.98 [+2.62,+5.20]	+1.99 [+1.47,+2.54]	+2.25 [+1.68,+2.84]	+3.16 [+2.64,+3.73]

F REPRESENTATION VISUALIZATIONS

F.1 FROZEN TRIAL-LEVEL PROTOCOL

We visualize frozen trial representations for all four EEG foundation backbones on the same twelve validation panels. Trial identities, labels, subject and stimulus metadata, and plotting settings are shared across backbones. $R0$ is the pretrained encoder readout and $S2$ is the final PACE readout. Features are flattened to $\mathbf{h}_i$ and L2-normalized to $\mathbf{z}_i$. The cosine similarity and equivalent unit-sphere distance are

$$\mathbf{z}_i = \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|_2}, \qquad K_{ij} = \mathbf{z}_i^\top \mathbf{z}_j, \qquad D_{ij} = \sqrt{\max(0, 2 - 2K_{ij})}. \tag{23}$$

Each panel uses Barnes–Hut t-SNE with perplexity 30, seed 0, early exaggeration 12, automatic learning rate, and at most 1,500 iterations. DEAP and MEEG use binary valence labels.

For the original-space neighborhood score, the eligible candidate set $\mathcal{C}_i$ excludes same-subject and known same-stimulus trials. The ten candidates with largest K_{ij} form $\mathcal{N}_{10}(i)$; the same-label kNN@10 score for trial i is

$$q_i = \frac{1}{10} \sum_{j \in \mathcal{N}_{10}(i)} \mathbf{1}[y_j = y_i]. \tag{24}$$

Here y_i is the emotion label of trial i. Scores are averaged within class–subject cells, equally over subjects within each class, and finally over classes. The random reference uses the same candidate sets and averaging. The t-SNE coordinates are descriptive layouts; their absolute positions are not compared across panels.

Table 22. Complete fixed-stage neighborhood statistics. Values are percentages in the original feature space; Random is the candidate-set reference, and each backbone is compared between $R0$ and final $S2$. The final column is $\Delta = S2 - R0$ in percentage points.

CBraMod and LaBraM

Backbone	Target	Random	$R0$	$S2$	Δ
CBraMod	SEED	28.57	44.07	64.81	+20.74
CBraMod	SEED-IV	23.94	32.87	45.97	+13.10
CBraMod	SEED-V	18.18	19.41	22.52	+3.11
CBraMod	SEED-VII	13.22	15.42	18.39	+2.98
CBraMod	FACED	7.82	7.80	8.51	+0.71
CBraMod	EAV	19.19	25.87	27.80	+1.93
CBraMod	MEEG	50.00	68.75	59.33	-9.42
CBraMod	MC-ima	9.95	9.57	10.54	+0.97
CBraMod	MC-vid	9.92	10.00	12.04	+2.03
CBraMod	ImaginedEmotion	6.67	7.52	8.57	+1.05
CBraMod	REFED	14.29	12.89	17.00	+4.11
CBraMod	DEAP	49.44	52.64	52.56	-0.08
LaBraM	SEED	28.57	42.15	67.78	+25.63
LaBraM	SEED-IV	23.94	30.37	40.83	+10.46
LaBraM	SEED-V	18.18	20.74	25.85	+5.11
LaBraM	SEED-VII	13.22	14.99	17.22	+2.23
LaBraM	FACED	7.82	8.39	9.64	+1.26
LaBraM	EAV	19.19	26.34	29.09	+2.75
LaBraM	MEEG	50.00	73.42	66.50	-6.92
LaBraM	MC-ima	9.95	9.42	8.51	-0.91
LaBraM	MC-vid	9.92	9.06	12.88	+3.81
LaBraM	ImaginedEmotion	6.67	7.90	7.71	-0.19
LaBraM	REFED	14.29	15.00	18.67	+3.67
LaBraM	DEAP	49.44	48.89	52.88	+3.99

CSBrain and REVE

Backbone	Target	Random	$R0$	$S2$	Δ
CSBrain	SEED	28.57	39.48	59.56	+20.08
CSBrain	SEED-IV	23.94	27.31	36.62	+9.31
CSBrain	SEED-V	18.18	19.41	21.48	+2.07
CSBrain	SEED-VII	13.22	15.22	19.35	+4.12
CSBrain	FACED	7.82	7.41	8.17	+0.76
CSBrain	EAV	19.19	25.37	24.50	-0.87
CSBrain	MEEG	50.00	64.67	63.00	-1.67
CSBrain	MC-ima	9.95	9.68	9.88	+0.20
CSBrain	MC-vid	9.92	9.87	9.71	-0.16
CSBrain	ImaginedEmotion	6.67	7.33	6.86	-0.48
CSBrain	REFED	14.29	15.00	14.89	-0.11
CSBrain	DEAP	49.44	48.74	54.61	+5.87
REVE	SEED	28.57	38.22	65.85	+27.63
REVE	SEED-IV	23.94	28.98	42.64	+13.66
REVE	SEED-V	18.18	16.81	24.81	+8.00
REVE	SEED-VII	13.22	13.35	20.37	+7.02
REVE	FACED	7.82	7.69	9.33	+1.63
REVE	EAV	19.19	25.47	30.30	+4.83
REVE	MEEG	50.00	70.33	67.50	-2.83
REVE	MC-ima	9.95	9.89	10.34	+0.46
REVE	MC-vid	9.92	10.19	15.13	+4.93
REVE	ImaginedEmotion	6.67	7.43	7.71	+0.29
REVE	REFED	14.29	13.67	20.11	+6.44
REVE	DEAP	49.44	50.14	53.88	+3.74

F.2 CBraMod

The complete CBraMod fixed-readout scores are reported in Table 22. PACE improves 10 of the 12 panels; the two decreases occur on MEEG and DEAP.

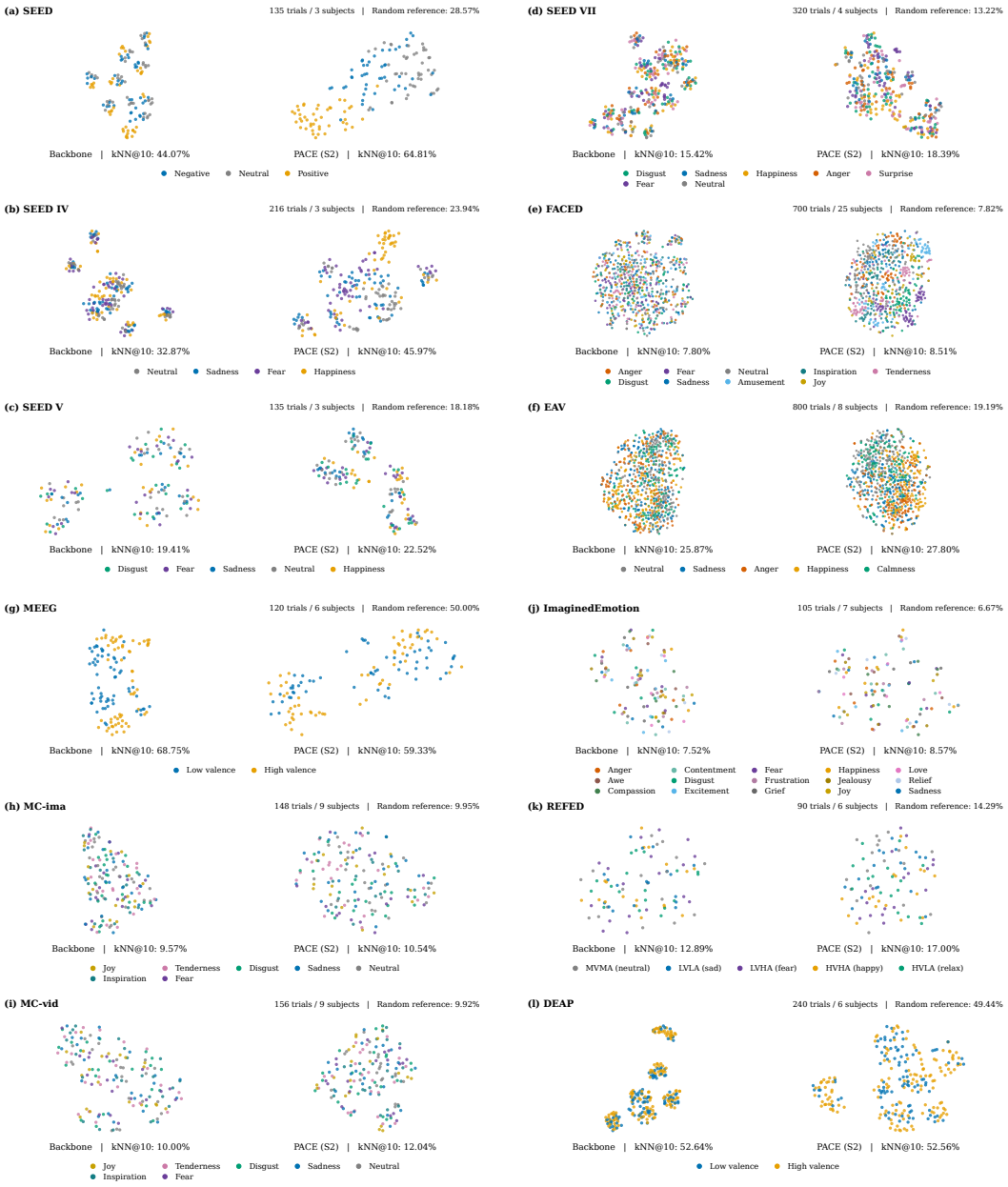

Figure 5. CBraMod trial-level t-SNE across all twelve diagnostic panels. The four tiles contain the same rows as the original per-page layout; each row compares the frozen backbone readout with the final PACE readout. Colors indicate native emotion labels, and values are same-emotion kNN@10 computed in the original feature space.

F.3 LABRAM

LaBraM shows fixed-readout improvements on 9 of the 12 panels, with the largest gains on SEED and SEED-IV.

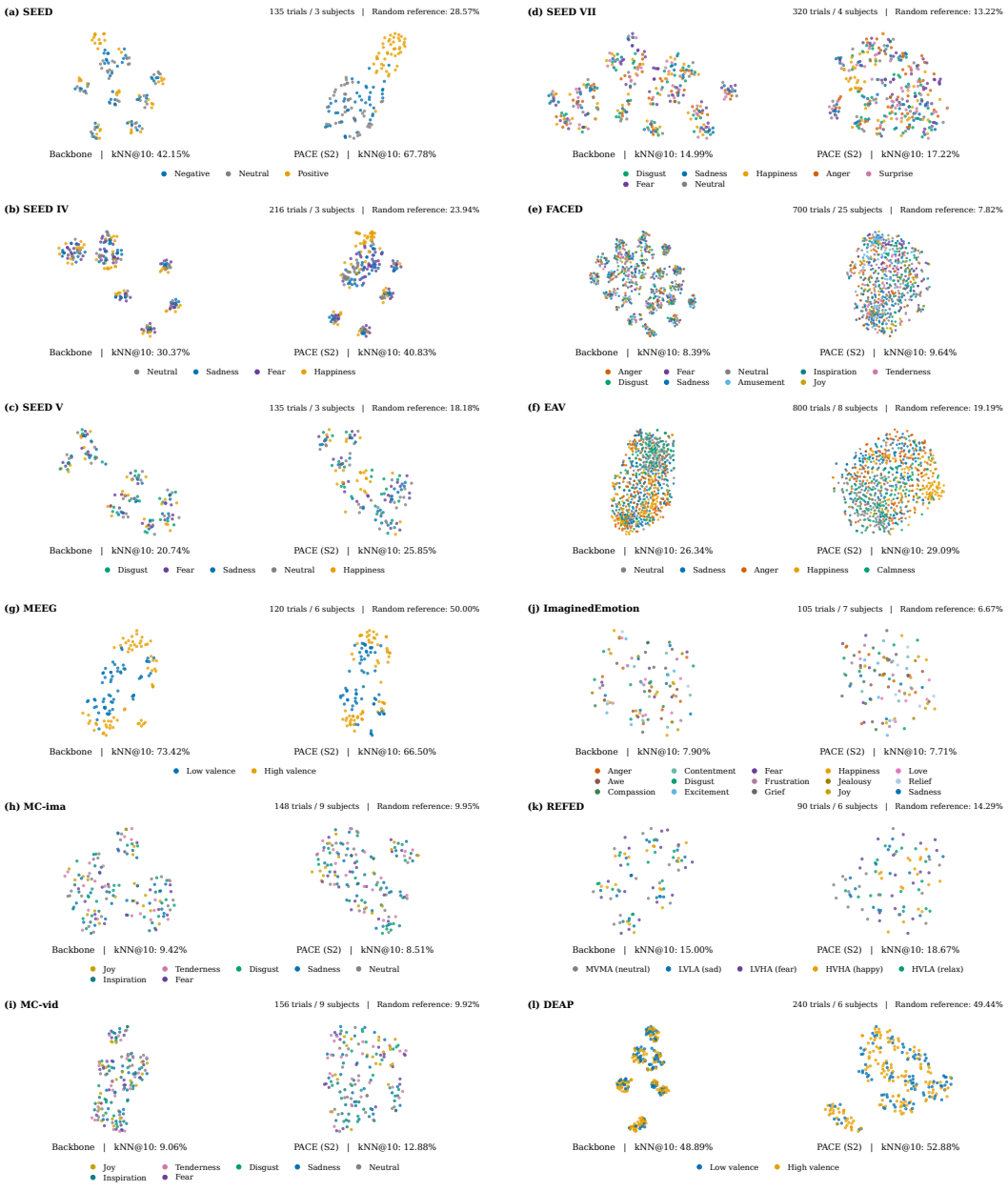

Figure 6. LaBraM trial-level t-SNE across all twelve diagnostic panels. The four tiles use the same validation trials and shared t-SNE settings as the other backbones.

F.4 CSBRAIN

For CSBrain, the fixed final readout improves 7 of 12 panels. The strongest gains are on the video datasets.

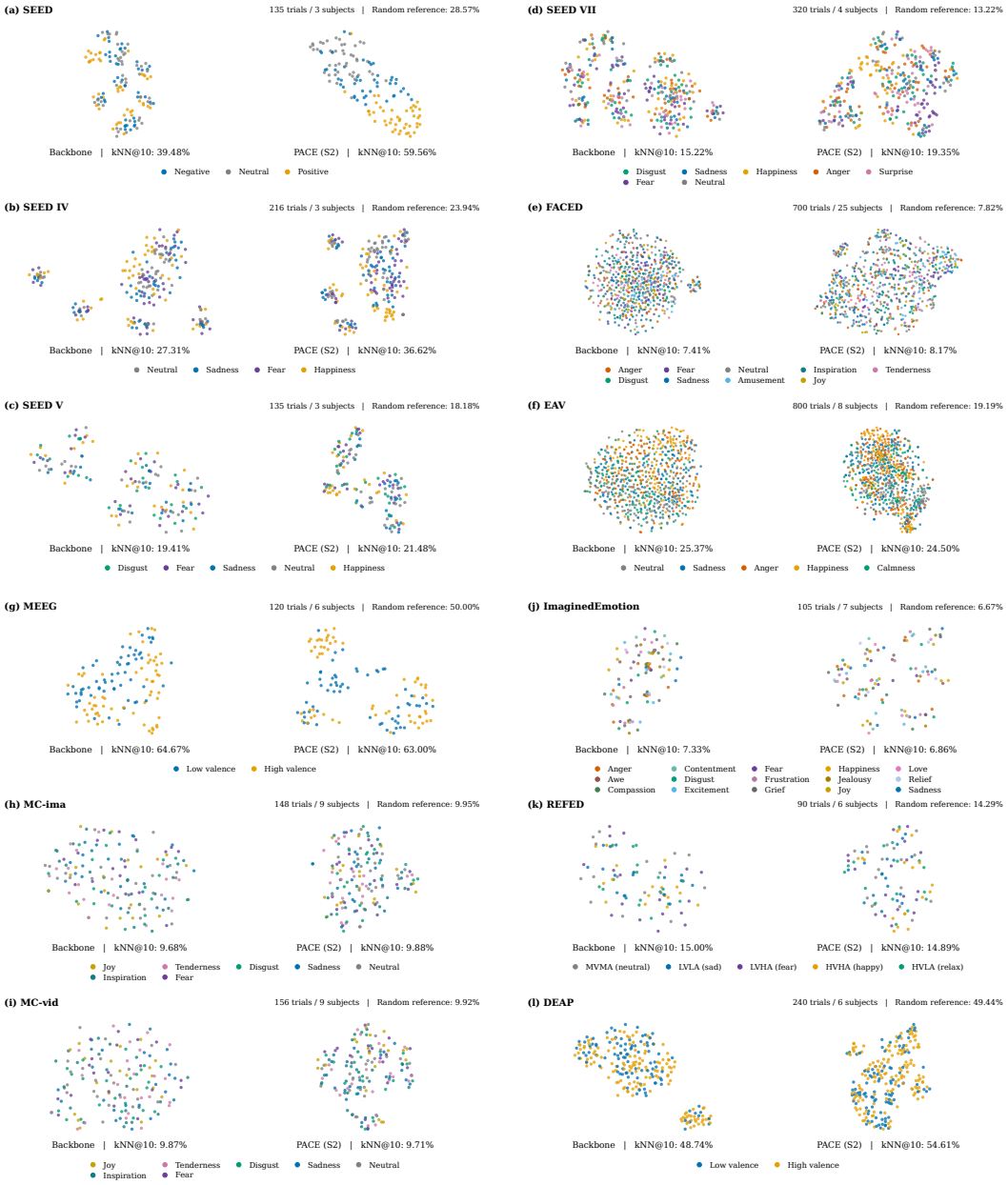

Figure 7. CSBrain trial-level t-SNE across all twelve diagnostic panels. Each tile reports the original-space neighborhood score below the corresponding pair of layouts.

F.5 REVE

REVE improves 11 of the 12 fixed-readout panels, with consistent gains across the video and imagery groups.

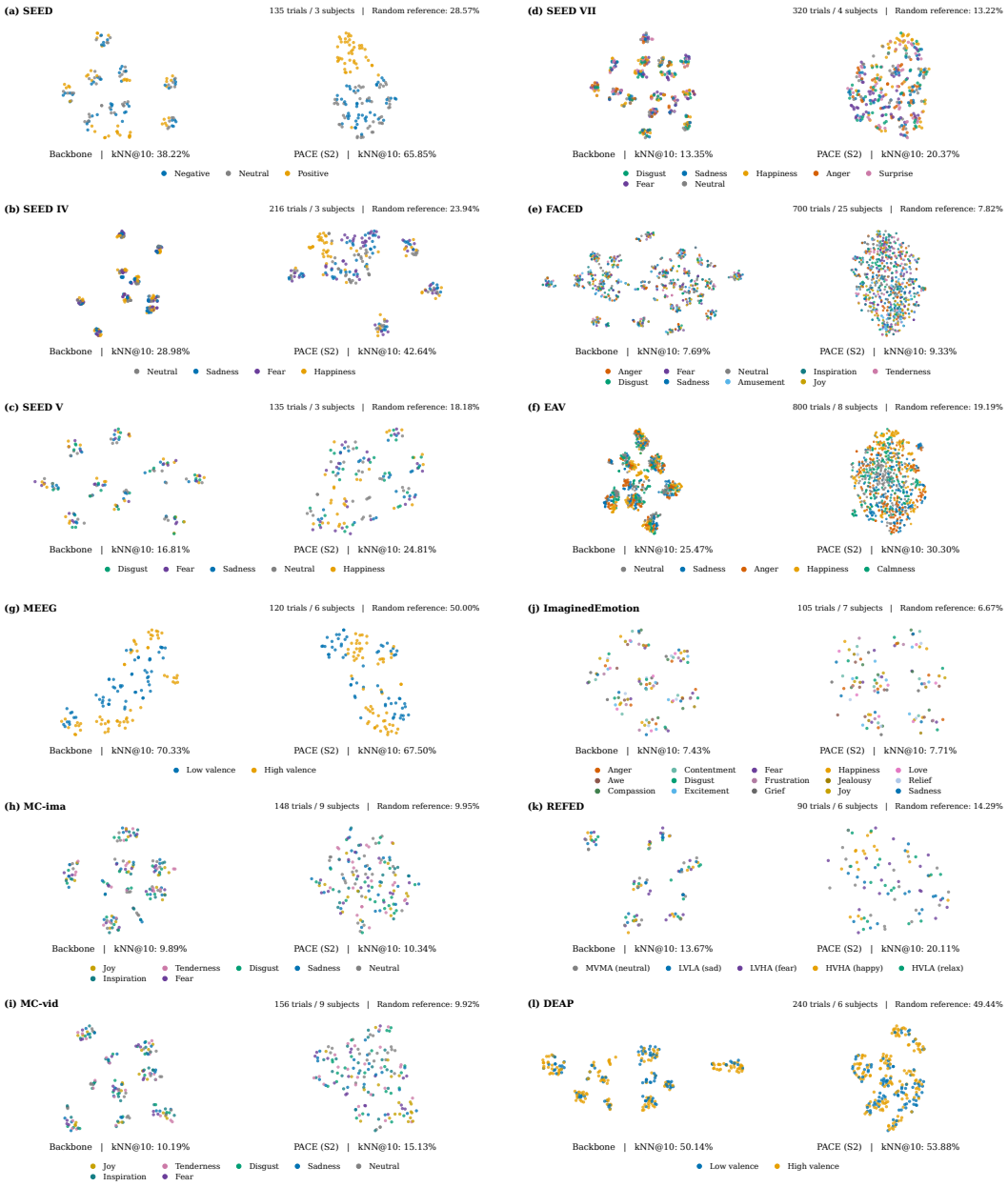

Figure 8. REVE trial-level t-SNE across all twelve diagnostic panels. Layouts use perplexity 30 and seed 0, with the same plotting protocol used for the other backbones.