UNIPER: TOWARDS A UNIFIED PREDICTIVE MODEL OF AFFECTIVE EEG

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG emotion recognition requires transferable representations, but general EEG pretraining often draws on clinical and other non-affective recordings. Learning transferable emotion representations from multiple affective datasets remains challenging because stimuli, participants, sensor layouts, and annotation schemes differ across sources. We introduce **UniPER**, a latent predictive framework that learns shared affective EEG representations without requiring source emotion labels or cross-dataset label alignment. UniPER trains a montage-aware online encoder to predict masked latent content from a full-view target encoder updated by an exponential moving average (EMA). **Position-matched content** subtracts each encoder's zero-waveform response at the same sensor and time location; a matched zero reference also centers the predictor output. Masked prediction and context identity learn from waveform content and source-local recording events, while within and cross-source relations organize latent similarities. We pretrain UniPER on nine affective EEG datasets and evaluate it on six held-out target datasets under subject-dependent, subject-independent, and zero-shot protocols, retaining each target's emotion labels. Under subject-dependent adaptation, UniPER achieves the **highest balanced accuracy on five of six targets**, reaching **65.60% on FACED**, **5.97 percentage points** above the strongest evaluated baseline. Component ablations examine zero references, local prediction, context identity, source relations, and representation regularization; their effects vary by target and metric. These results support multi-source affective pretraining through latent prediction as a route to reusable emotion representations across heterogeneous EEG datasets.

1 INTRODUCTION

Electroencephalography (EEG) offers a non-invasive, time-resolved view of brain activity, making EEG emotion recognition promising for affect-aware interfaces. The task is difficult because emotion-related patterns are embedded in noisy, non-stationary signals and vary across people and recording sessions (Shen et al., 2023; Li et al., 2022). Across datasets, stimuli, acquisition setups, and emotion annotations also change, as illustrated by SEED, DEAP, FACED, and MPED (Zheng & Lu, 2015; Koelstra et al., 2012; Chen et al., 2023; Song et al., 2019). A decoder can therefore succeed within one protocol by relying on cues that fail to transfer.

Self-supervised learning has improved the transfer of EEG emotion representations between participants. CLISA aligns responses to the same stimulus across subjects, and MV-SSTMA learns from masked spectral, spatial, and temporal views (Shen et al., 2023; Li et al., 2022). These methods reduce dependence on emotion labels, but a new participant within one corpus usually shares its stimuli, recording setup, and label vocabulary. A new dataset can change all three. E2STN and the multi-source masked pretraining of MMM extend the question beyond a single corpus (Zhou et al., 2025; Yi et al., 2023), motivating a closer study of representations learned across affective datasets.

Pretraining on multiple affective sources exposes an encoder to these changes before adaptation to an unseen dataset. General EEG foundation models often draw on clinical or mixed-task recordings without centering affective EEG (Kostas et al., 2021; Yang et al., 2023a; Jiang et al., 2024; Wang et al., 2025; El Ouahidi et al., 2025); NeurIPT also includes affective data, and MMM combines emotion datasets (Fang et al., 2025; Yi et al., 2023). More directly, mdJPT learns from multiple emotion corpora without emotion labels using covariance alignment and temporally paired inter-

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.

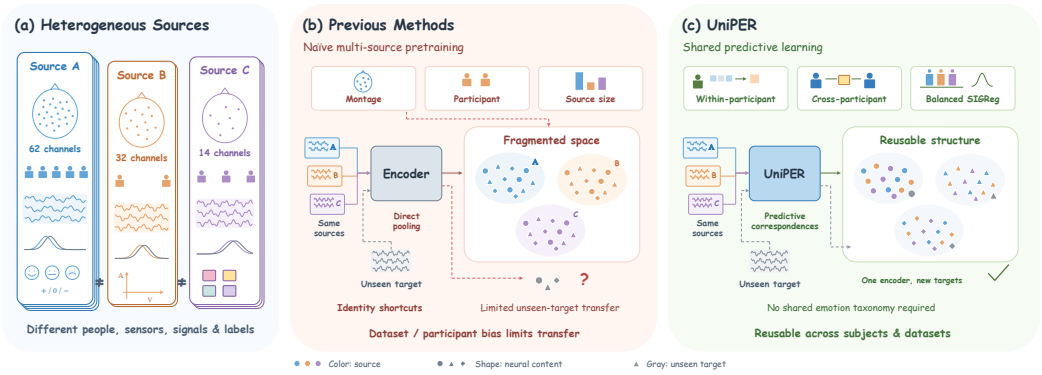

Figure 1: **Motivating paradigm.** (a) Three illustrative EEG sources differ in montage, participant and signal properties, and labels. (b) A schematic of source-separated structure after naive pooling. (c) The intended UniPER representation shares predictive structure while retaining distinct content groups for an unseen target. Colors mark sources, shapes mark illustrative content, and gray marks the unseen target. All point patterns are conceptual illustrations.

subject views, whereas EMOD aligns corpora through valence–arousal annotations (Zhang et al., 2025; Chen et al., 2026). Across sources, matched events and compatible emotion labels may be unavailable. The training signal must therefore remain meaningful when stimuli and label spaces differ, as the schematic in Figure 1 illustrates.

Latent predictive learning offers a signal that can be defined within each recording: visible EEG context predicts the representation of hidden content. Data2vec and I-JEPA established this approach in other modalities, while EEG2Rep and EEGPT brought masked representation learning to EEG (Baevski et al., 2022; Assran et al., 2023; Mohammadi Foumani et al., 2024; Wang et al., 2024). MGCRL has also coupled JEPA-style prediction with contrastive learning for transfer from FACED to SEED-series tasks (Weng et al., 2026). For heterogeneous affective sources, the central design question is how to define the predicted content and the window representation retained for downstream emotion tasks.

Our setting is to pretrain one encoder on heterogeneous affective EEG sources without using their emotion labels, then transfer it to unseen datasets that retain their own label spaces. To this end, we propose UniPER, a multi-source latent predictive framework (Figure 2). A masked online encoder predicts content from a full-view exponential-moving-average teacher. At each sensor–time position, content is defined relative to the encoder's zero-waveform response; the predictor also subtracts its response to zero memory under the same queries. Context identity and source relations organize predictive similarities, while distributional regularization acts on the pooled online content used downstream. Our main contributions are:

- **Multi-source affective EEG pretraining.** We train one montage-aware encoder across heterogeneous emotion recordings without source emotion labels or a shared emotion taxonomy. The encoder exports a common representation that can be adapted to each target's native task.
- **Position-matched latent prediction.** We define encoder content and predictor outputs using matched zero-input references. Masked prediction, context identity, and source relations supervise predictive learning; source-centered SIGReg regularizes the exported window representation.
- **Transfer across held-out emotion datasets.** We evaluate six unseen targets under subject-dependent adaptation, subject-independent transfer, and zero-shot inference. UniPER achieves the highest mean balanced accuracy on five of six targets under subject-dependent adaptation.

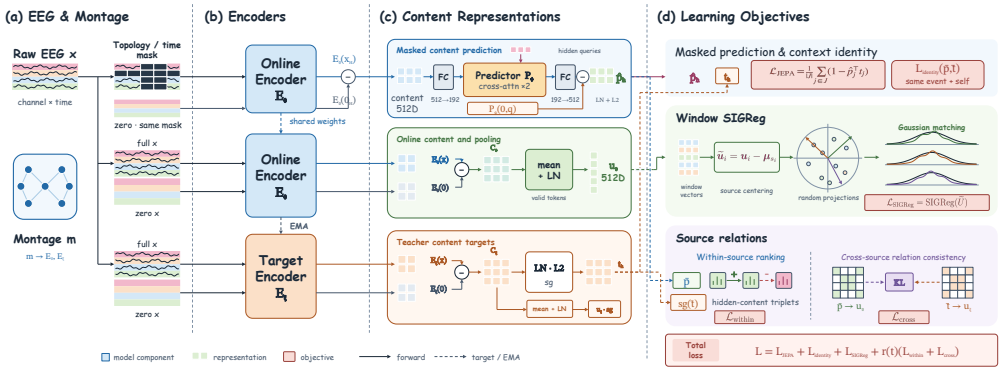

Figure 2: **Model architecture and objectives.** Matched zero-waveform passes define masked online, full online, and full teacher content. The predictor subtracts its zero-memory response under the same queries. Masked prediction and context identity supervise hidden content; within-source ranking uses aggregated hidden predictions and teacher targets. Cross-source relations compare predictive queries against full-window endpoint keys. SIGReg acts on the source-centered online endpoint.

2 METHODS

UniPER combines a montage-aware online encoder, an exponential-moving-average (EMA) teacher, and a masked-content predictor (Figure 2). Position-matched references define the content learned by the prediction and relation objectives. Source-centered regularization acts on the window representation exported for transfer.

2.1 MULTI-SOURCE INPUTS AND ENCODER

Let $\mathcal{S} = \{\mathcal{D}_s\}_{s=1}^{S}$ denote the source corpora. A preprocessed EEG window $X \in \mathbb{R}^{C\times 625}$ contains five seconds at 125 Hz, with electrode coordinates $m \in \mathbb{R}^{C\times 3}$ and a channel-validity mask v. Each recording also has a participant identifier b and a source-local event key $e = (\text{session}, \text{stimulus}, \text{window})$. Source learning uses the waveform and these recording identifiers; downstream readouts use the target's native emotion labels.

Each channel is divided into 25 non-overlapping patches of 25 samples. Parallel patch-local convolutions with kernels 3, 7, and 15 and eight filters each feed a linear projection to 192D. A linear electrode-coordinate embedding and sinusoidal time encoding locate each token. Three spatial-then-temporal attention blocks operate at 192D with six heads. A $192 \to 512$ projection and GELU precede one 512D, eight-head spatiotemporal block and a $512 \to 512$ readout. Thus, the token path is $[B, C, 625] \to [B, C, 25, 25] \to [B, C, 25, 192] \to [B, C, 25, 512]$. The validity mask selects the observed sensor–time positions V. Shared parameters and electrode geometry provide a common interface across montages (Yi et al., 2023; Wang et al., 2024).

2.2 POSITION-MATCHED CONTENT AND EXPORTED REPRESENTATION

For $\eta \in \{\theta, \xi\}$, let E_η denote the online or teacher encoder. Content is the response relative to a zero waveform at the same sensor–time position:

$$c_j^\eta(X; m, v) = E_\eta(X; m, v)_j - E_\eta(0; m, v)_j, \qquad j \in V. \tag{1}$$

Here 0 is the zero waveform in encoder-input coordinates. The real and reference passes share electrode positions, temporal positions, and validity conditions. The masked online path uses the same construction with matching visibility in both passes. The complete online and teacher paths use all valid positions.

The complete-input online content supplies the downstream representation:

$$u_\theta(X; m, v) = \mathrm{LN}\left(\frac{1}{|V|}\sum_{j\in V} c_j^\theta(X; m, v)\right) \in \mathbb{R}^{512}. \tag{2}$$

The teacher endpoint u_ξ uses the same pooling rule with stopped gradients. This endpoint and the hidden-content summaries below serve distinct roles in training.

2.3 Masked prediction and context identity

A neighboring-electrode cluster mask and a contiguous temporal mask each cover approximately 28% of their axis. Their union defines hidden valid positions M and visible context $H = V \setminus M$. The recorded hidden fraction is approximately 47%. The full-view teacher supplies hidden-position targets, following latent predictive learning (Baevski et al., 2022; Assran et al., 2023; Mohammadi Foumani et al., 2024; Wang et al., 2024).

The predictor P_ψ projects visible 512D content to 192D memory, applies two six-head cross-attention blocks to hidden-position queries q_j, and projects back to 512D. It evaluates a matched reference using the same queries and zero memory. With $\nu(a) = a/\|a\|_2$, its output and the teacher target are

$$\begin{aligned} \widehat{p}_{ij} &= \nu\big(\mathrm{LN}\big[P_\psi(c^\theta_{i,H}, q_j) - P_\psi(0, q_j)\big]\big)\,, \\ t_{ij} &= \mathrm{sg}\Big[\nu\Big(\mathrm{LN}(c^\xi_{ij})\Big)\Big]\,. \end{aligned} \tag{3}$$

The zero memory has the same shape as the real memory. Let $J_i \subseteq M_i$ be the hidden positions evaluated for window i; the predictor processes at most 128 hidden queries at a time. Position-wise cosine prediction and the two window summaries are

$$\mathcal{L}_{\mathrm{JEPA}} = \operatorname*{mean}_i \frac{1}{|J_i|}\sum_{j\in J_i}(1 - \widehat{p}_{ij}^\top t_{ij}), \quad \bar{p}_i = \nu\left(\frac{1}{|J_i|}\sum_{j\in J_i}\widehat{p}_{ij}\right), \quad \bar{t}_i = \mathrm{sg}\left[\nu\left(\frac{1}{|J_i|}\sum_{j\in J_i} t_{ij}\right)\right]. \tag{4}$$

Context identity matches $\bar{p}_i$ to hidden teacher summaries within the same source batch B_s. Define $P_i = \{j \in B_s : e_j = e_i\}$ and temperature $\tau_c = 0.20$. The multi-positive objective is

$$\mathcal{L}_{\mathrm{identity}} = \operatorname*{mean}_s \frac{1}{|B_s|}\sum_{i\in B_s} -\log\frac{\sum_{j\in P_i}\exp(\bar{p}_i^\top \bar{t}_j/\tau_c)}{\sum_{j\in B_s}\exp(\bar{p}_i^\top \bar{t}_j/\tau_c)}. \tag{5}$$

The positive set includes the window's own teacher pair and any other window with the same event key, including matching events from other participants. Each source batch gives a 12×12 similarity matrix. The numerator pools positive probability mass, and the denominator includes all candidate events.

2.4 Within- and cross-source relations

Within-source ranking. Five pairable sources additionally provide event-matched participant triplets. For sampled $(i, j, k) \in \mathcal{T}_s$, the anchor i and positive j have different participants and identical event keys. The negative k has the same participant, session, and window index as j, with a different stimulus. Using $\tau_r = 0.20$,

$$\ell_s = \begin{cases} \dfrac{1}{|\mathcal{T}_s|}\displaystyle\sum_{(i,j,k)\in\mathcal{T}_s} \mathrm{softplus}\left(\frac{\bar{p}_i^\top \bar{t}_k - \bar{p}_i^\top \bar{t}_j}{\tau_r}\right), & |\mathcal{T}_s| > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad \mathcal{L}_{\mathrm{within}} = \operatorname*{mean}_s \ell_s. \tag{6}$$

Sources with this relation disabled contribute zero to the current source-batch average. The ranking operates on aggregated hidden predictions and teacher targets.

Cross-source similarity distributions. Predictive summaries act as queries and complete-window endpoints as keys. For each ordered pair of sources (a, b) present in an update, define, for $i \in B_a$ and $j \in B_b$,

$$Q_\theta^{a\to b}(j \mid i) = \frac{\exp(\bar{p}_i^\top \nu(u_{\theta,j})/\tau_r)}{\sum_{k\in B_b} \exp(\bar{p}_i^\top \nu(u_{\theta,k})/\tau_r)}, \qquad Q_\xi^{a\to b}(j \mid i) = \frac{\exp(\bar{t}_i^\top \nu(u_{\xi,j})/\tau_r)}{\sum_{k\in B_b} \exp(\bar{t}_i^\top \nu(u_{\xi,k})/\tau_r)}. \tag{7}$$

All four vector types are L2-normalized for cosine similarity. Softmax runs over the other source's candidate windows. With $\mathcal{O}$ the ordered source pairs, the loss is

$$\mathcal{L}_{\text{cross}} = \frac{1}{|\mathcal{O}|} \sum_{(a,b)\in\mathcal{O}} \frac{1}{|B_a|} \sum_{i\in B_a} D_{\text{KL}}\big(\text{sg}[Q_\xi^{a\to b}(\cdot \mid i)] \,\|\, Q_\theta^{a\to b}(\cdot \mid i)\big)\,. \tag{8}$$

The two source batches in each update supply the vectors for these distributions, yielding two comparison directions, weighted equally. The objective transfers the teacher's relative similarity structure across source boundaries.

2.5 Endpoint regularization and optimization

SIGReg (Balestriero & LeCun, 2025) regularizes the complete-input online endpoint. Source centering gives $\widetilde{u}_i = u_{\theta,i} - \mu_{s(i)}$, where μ_s is the mean endpoint for source s in the update. For unit projection direction a_k, the empirical characteristic function and Gaussian reference are

$$\widehat{\varphi}_k(\omega) = \frac{1}{N} \sum_{i=1}^{N} e^{\mathrm{i}\omega a_k^\top \widetilde{u}_i}, \qquad \varphi_0(\omega) = e^{-\omega^2/2}. \tag{9}$$

The loss averages the weighted squared discrepancy $|\widehat{\varphi}_k(\omega) - \varphi_0(\omega)|^2$ over 256 directions and 17 integration points on $[0, 3]$. The update contains 24 windows from two source batches. Centering is part of the regularizer; transfer uses Eq. 2.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{JEPA}} + \mathcal{L}_{\text{identity}} + \mathcal{L}_{\text{SIGReg}} + r(t)\big(\mathcal{L}_{\text{within}} + \mathcal{L}_{\text{cross}}\big). \tag{10}$$

For training progress $t \in [0, 1]$, $r(t)$ is zero through 20%, increases linearly to 0.05 at 40%, and remains 0.05 thereafter. Gradient updates optimize the online encoder and predictor. The teacher starts as an online-encoder copy and follows each optimizer step by $\xi \leftarrow 0.996\,\xi + 0.004\,\theta$. Downstream inference exports the complete-input online endpoint. Appendix C gives the numerical configuration.

3 Experiments

3.1 Evaluation setting

We evaluate transfer to six held-out native-label emotion tasks: SEED (three classes), SEED-IV (four), SEED-V (five), DEAP (binary valence), FACED (nine), and MPED (seven). The source corpus contains 77,850 retained windows from nine affective datasets; the six target caches contain 97,985 windows (Table 1). Source pretraining uses EEG and recording-event metadata, with the six target datasets excluded. Each downstream task retains its own class vocabulary and readout.

The source recordings span 8–128 measured electrodes, while the targets use 32 or 62. Five-second windows are resampled to 125 Hz and represented with electrode coordinates and a validity mask. Subject-dependent (SD) adaptation separates complete trials within each participant, using approximately 75% for training and 25% for testing. Three fixed split seeds define the partitions. Subject-independent (SI) evaluation uses three separate participant-disjoint 75/25 partitions. We also evaluate zero-shot transfer under the protocol in Appendix E.1. Appendix E describes the data organization; Appendix C gives the configuration. Source pretraining used two NVIDIA H100 GPUs on HPC3; target adaptation and evaluation used one H100 GPU per run.

We compare supervised EEGNet, EEG Conformer, TSception, DGCNN, and RGNN (Lawhern et al., 2018; Song et al., 2023; Ding et al., 2023; Song et al., 2020; Zhong et al., 2022); self-supervised CLISA, ContraWR, and MMM (Shen et al., 2023; Yang et al., 2023b; Yi et al., 2023); and pretrained

Table 1: Six held-out targets and nine pretraining sources. Rates and channels describe the analyzed recordings before common resampling. Window counts describe the retained 5-s source corpus and target caches; subject counts give original dataset sizes. The FACED cache contains 51 of 123 participants. Label lists native target classes or source classes/rating fields. Pretraining uses EEG and recording-event metadata.

Dataset	Rate (Hz)	#Channels	Window	#Windows	#Subjects	Label
Held-out target caches: 97,985 windows						
SEED	1000	62	5 s	30,375	15	3
SEED-IV	1000	62	5 s	4,946	15	4
SEED-V	1000	62	5 s	23,280	16	5
DEAP	512	32	5 s	15,360	32	2
FACED	250	32	5 s	8,568	123	9
MPED	1000	62	5 s	15,456	23	7
Pretraining sources: 77,850 windows						
DREAMER	128	14	5 s	17,020	23	3
EmoEEG-MC	600, 1200	64	5 s	11,424	60	7
MDME	300	21	5 s	11,824	73	3
SEED-VII	1000	62	5 s	6,014	20	7
DENS	250	128	5 s	7,496	40	7
EEGEmotions-27	256	14	5 s	15,670	88	27
GAMEEMO	128	14	5 s	6,844	28	4
GU	128	32	5 s	180	10	3
LUMED-2	500	8	5 s	1,378	13	3

BENDR, BIOT, LaBraM, CBraMod, mdJPT, and EMOD (Kostas et al., 2021; Yang et al., 2023a; Jiang et al., 2024; Wang et al., 2025; Zhang et al., 2025; Chen et al., 2026). These families cover supervised task learning, self-supervised representations, and transfer from pretrained models. Metrics are balanced accuracy (BACC, mean class recall), macro-F1, and macro one-versus-rest AUROC. We compare methods within each target's native task; the class-uniform BACC reference ranges from 50% on DEAP to 11.1% on FACED.

3.2 Transfer to unseen affective datasets

Table 2 reports SD adaptation. UniPER achieves the highest mean BACC on five of six targets. On FACED, it reaches 65.60%, exceeding EMOD's 59.63% by 5.97 percentage points. On SEED, it obtains 84.33% BACC and 84.23% macro-F1; the next-highest BACC is mdJPT's 80.85%. MPED shows a smaller numerical difference, 65.22% versus EMOD's 64.77%. SEED-IV favors other methods: UniPER scores 74.34%, compared with 75.55% for EMOD, 74.96% for mdJPT, and 74.73% for CBraMod. These results characterize transfer across distinct acquisition settings and native emotion taxonomies.

SI evaluation tests transfer to participants outside target training. UniPER's SI BACC ranges from 44.06% on FACED to 77.54% on DEAP (Appendix F). Across the six targets, SD scores exceed SI scores, with gaps from 9.83 points on DEAP to 21.54 on FACED. The ordering highlights participant variability alongside the dataset shift addressed by source pretraining. SEED-IV also favors the comparison methods under SI and zero-shot inference, showing that method rankings depend on the target and readout protocol.

Figure 3 presents class-wise error patterns on SEED-V, complementing the aggregate BACC, macro-F1, and AUROC results.

3.3 Zero-shot and label-limited transfer

Table 3 evaluates the pretrained encoder under the zero-shot protocol. UniPER obtains 22.85% BACC on FACED and 30.46% on MPED, above their respective class-uniform references of 11.1% and 14.3%. On DEAP, UniPER has the highest zero-shot BACC, 70.73%, while mdJPT has higher AUROC, 75.13% versus 71.01%. This distinction separates the evaluated classification decision from ranking performance.

The zero-shot and adapted protocols address complementary uses of the representation. End-to-end adaptation learns a target-specific linear classifier together with the encoder. The exported representation is the mean of valid complete-input online content tokens followed by LayerNorm.

Table 2: Subject-dependent results on the six target tasks. The first and second blocks show SEED/SEED-IV/SEED-V and DEAP/FACED/MPED, respectively. Cells give mean ± standard deviation percentages for BACC, macro-F1, and macro one-versus-rest AUROC. Bold and underline mark the two highest means in each column.

Method	SEED			SEED-IV			SEED-V		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	71.09±1.42	70.70±1.62	79.74±1.06	59.67±1.79	58.68±1.56	80.00±1.52	50.06±1.26	49.09±1.09	72.70±1.15
EEG Conformer	57.30±4.34	56.57±5.75	72.69±2.22	63.37±1.10	62.26±1.05	82.88±0.84	55.20±1.36	54.31±1.15	74.54±1.03
TSception	53.02±1.73	52.72±1.96	68.19±1.74	65.79±1.33	65.33±1.51	84.46±0.80	56.55±1.55	55.98±1.63	75.74±1.41
DGCNN	76.68±0.82	76.05±0.57	82.30±0.47	68.44±1.76	67.64±1.79	85.84±1.30	59.75±1.23	58.94±1.10	76.99±0.95
RGNN	40.15±1.67	37.56±4.16	55.27±1.46	69.63±1.54	68.96±1.39	86.75±1.41	60.00±1.54	58.88±1.74	77.59±1.11
SSL									
CLISA	78.76±1.01	78.13±0.99	84.46±0.68	72.04±1.91	71.09±2.00	88.28±1.63	62.89±1.45	62.29±1.59	78.89±0.97
ContraWR	78.19±1.05	77.66±1.03	84.27±0.72	70.21±1.34	69.42±1.34	87.00±1.12	61.62±1.43	60.82±1.54	78.48±1.24
MMM	79.31±0.72	78.69±0.87	84.84±0.48	72.48±1.90	71.59±2.01	88.49±1.68	63.87±1.12	63.12±1.05	79.85±1.06
Pretraining									
BENDR	76.23±1.23	75.91±1.02	82.93±0.84	68.16±1.20	67.09±1.19	85.47±1.00	58.97±1.59	57.89±1.86	76.45±1.18
BIOT	78.52±0.68	77.86±0.86	83.45±0.39	70.60±1.75	69.44±1.60	86.16±1.45	62.18±1.13	61.35±1.20	79.43±1.01
LaBraM	80.25±1.34	79.39±1.15	84.60±0.89	73.06±1.18	72.00±1.16	87.68±0.76	64.35±1.79	63.60±1.64	79.64±1.24
CBraMod	79.35±0.66	78.71±0.54	84.50±0.39	74.73±1.87	73.56±1.65	**89.21**±1.65	65.84±1.52	64.85±1.59	80.33±1.20
EMOD	80.80±0.83	80.36±0.73	84.99±0.49	**75.55**±1.85	**74.79**±1.78	88.80±1.66	65.84±1.62	65.65±1.59	81.14±1.57
mdJPT	80.85±1.23	80.43±1.42	84.34±0.73	74.96±1.06	73.73±1.29	89.02±0.77	65.96±1.46	65.80±1.68	81.50±0.95
UniPER	**84.33**±1.96	**84.23**±1.51	**87.95**±1.30	74.34±1.33	74.09±1.45	88.08±0.53	**71.46**±1.17	**70.21**±1.46	**84.01**±0.99

Method	DEAP			FACED			MPED		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	72.41±1.55	71.75±1.41	80.31±1.07	43.55±1.23	42.46±1.02	72.53±1.00	46.83±1.14	45.55±1.02	73.99±0.86
EEG Conformer	75.13±2.00	74.64±1.97	81.96±1.81	11.95±1.70	3.47±1.62	58.97±1.68	51.28±1.65	49.99±1.43	76.50±1.08
TSception	52.83±0.65	52.70±0.62	53.89±0.56	27.77±0.65	27.07±0.95	67.31±0.51	53.95±1.82	52.70±2.00	77.05±1.67
DGCNN	77.71±1.49	77.35±1.50	83.91±0.95	10.53±0.07	2.63±0.31	51.66±1.57	56.59±1.30	55.51±1.45	78.57±0.85
RGNN	78.57±1.27	78.14±1.44	86.03±0.86	52.06±1.63	50.39±1.83	77.76±1.16	57.30±1.83	56.36±1.75	79.83±1.76
SSL									
CLISA	80.41±1.41	80.23±1.63	86.53±0.86	54.02±1.48	52.53±1.60	78.46±1.09	59.99±1.26	58.98±1.19	81.09±0.95
ContraWR	79.43±1.12	78.98±1.15	86.70±0.60	53.43±1.86	51.83±1.64	79.06±1.78	58.85±1.57	57.80±1.50	80.91±1.02
MMM	82.19±1.78	82.08±1.73	89.16±1.32	55.39±1.25	53.92±1.02	78.86±1.11	61.66±1.81	60.42±1.73	80.89±1.58
Pretraining									
BENDR	79.09±1.66	78.88±1.80	85.02±1.26	50.86±1.15	49.43±0.96	76.50±0.98	56.96±1.80	56.13±1.72	79.88±1.63
BIOT	80.97±1.88	80.47±2.12	87.62±1.31	53.81±1.37	52.38±1.46	78.72±1.14	59.81±1.47	58.77±1.30	80.36±1.11
LaBraM	82.74±1.12	81.94±0.96	89.05±1.00	58.08±1.81	56.66±1.84	80.53±1.68	61.79±1.83	60.71±2.05	81.08±1.36
CBraMod	83.79±1.15	83.00±1.40	89.13±1.10	57.32±1.46	54.63±1.62	79.13±1.20	63.49±1.70	62.50±1.69	82.20±1.29
EMOD	83.94±1.45	83.40±1.74	89.03±1.30	59.63±1.90	58.52±1.81	80.95±1.44	64.77±1.69	63.42±1.51	83.26±1.24
mdJPT	84.92±1.43	84.20±1.46	90.52±0.95	58.43±1.38	56.90±1.31	80.09±1.09	64.02±1.64	62.66±1.90	83.36±1.51
UniPER	**87.37**±1.58	**87.17**±1.75	**92.97**±1.26	**65.60**±1.14	**65.82**±1.19	**87.77**±1.31	**65.22**±1.23	**65.74**±1.24	**85.34**±1.29

Figure 4 summarizes performance as the labeled target-training trial budget increases. Each target contributes equally to the mean, retaining its native label space. The bars describe the relationship between target supervision and transfer performance across the evaluated budgets.

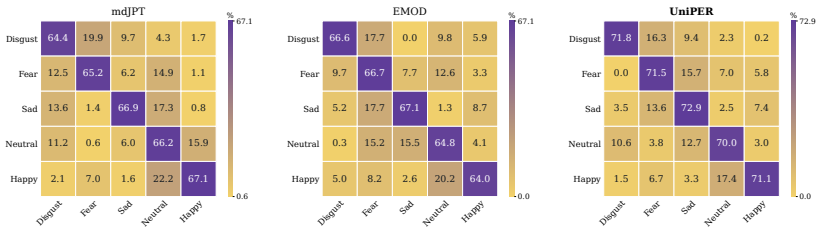

Figure 3: **SEED-V confusion matrices.** Panels show mdJPT, EMOD, and UniPER from left to right. Rows denote true classes and columns predicted classes. Cells are row-normalized percentages; each panel uses its observed color range.

Table 3: Zero-shot results on the six target tasks. The top block shows SEED/SEED-IV/SEED-V and the bottom block DEAP/FACED/MPED. Cells give mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC. Bold and underline mark the two highest means in each column.

Method	SEED			SEED-IV			SEED-V		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
UniPER	**58.29**±1.28	**57.19**±1.47	**75.54**±1.61	<u>45.28</u>±1.40	<u>44.15</u>±1.74	<u>68.57</u>±1.83	**45.87**±1.35	**44.87**±1.31	**66.39**±1.57
mdJPT	<u>56.29</u>±1.35	<u>55.21</u>±1.19	<u>68.79</u>±0.87	**46.10**±2.24	**44.52**±2.42	**70.08**±2.05	<u>42.60</u>±1.32	<u>41.74</u>±1.19	<u>65.57</u>±1.19
EMOD	51.01±1.41	50.18±1.21	65.21±1.14	43.17±1.66	41.89±1.62	67.93±1.32	37.88±1.98	36.58±2.20	62.27±1.76
LaBraM	42.24±1.62	41.08±1.70	58.34±1.45	40.26±1.37	39.04±1.67	65.62±0.84	32.16±1.45	30.79±1.50	59.22±1.36
CBraMod	47.12±1.59	46.47±1.59	61.99±1.36	38.11±1.49	36.89±1.61	63.69±1.06	34.73±1.90	33.85±2.00	60.65±1.49

Method	DEAP			FACED			MPED		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
UniPER	**70.73**±1.59	**70.68**±1.68	71.01±1.25	**22.85**±1.48	**22.26**±1.24	**63.75**±1.46	**30.46**±1.24	**30.57**±1.77	**67.16**±1.15
mdJPT	<u>68.28</u>±1.59	<u>67.79</u>±1.63	**75.13**±1.32	15.39±1.90	12.88±1.75	54.04±1.68	28.82±1.82	26.79±1.84	<u>62.06</u>±1.46
EMOD	66.69±2.22	66.11±2.43	<u>73.24</u>±1.98	<u>18.72</u>±1.85	<u>16.66</u>±1.68	<u>57.00</u>±1.68	<u>30.15</u>±1.76	<u>27.87</u>±1.48	61.93±1.31
LaBraM	62.70±1.49	61.64±1.40	69.73±1.21	16.79±1.58	12.21±1.35	53.61±1.05	24.15±1.72	21.74±1.47	58.60±1.31
CBraMod	64.31±1.97	63.33±1.72	70.54±1.78	17.36±1.91	14.39±1.67	54.33±1.66	27.31±2.16	25.62±2.12	60.38±1.82

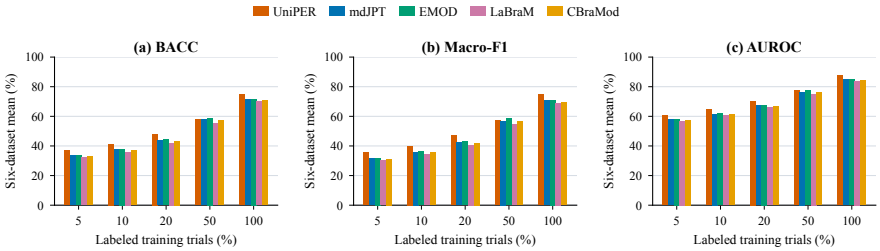

Figure 4: **Label-budget comparison.** Bars show the equal-target mean BACC, macro-F1, and AUROC across six targets for UniPER, mdJPT, EMOD, LaBraM, and CBraMod at 5, 10, 20, 50, and 100% of labeled target-training trials. The 100% LaBraM and CBraMod bars use the updated full-budget results. All panels use a 0–100% vertical scale.

Figure 5 provides a qualitative view of raw signals and pretrained representations. Each column uses its own target label vocabulary; the quantitative transfer comparison is given by the corresponding readout results.

3.4 Contribution of the training components

Table 4 compares training variants on SEED-IV, DEAP, and FACED. Removing zero references lowers BACC by 10.50, 13.53, and 4.56 points, respectively, while removing context identity lowers it by 8.02, 7.93, and 6.94 points. The corresponding decreases without local prediction are 11.48, 9.35, and 10.17 points. Removing the within-source relation lowers BACC by 2.61, 1.98, and 2.13 points, although FACED macro-F1 rises by 1.00 point. These effects vary by target and metric.

The relation controls further distinguish the use of event identities from the choice of relation objective. Shuffling event pairs reduces BACC by 8.23, 7.26, and 8.69 points on SEED-IV, DEAP, and FACED, larger decreases than removing the within-source relation. This pattern is consistent with incorrect event associations being more harmful than omitting that supervision. Removing the cross-source relation produces smaller BACC decreases of 2.59, 2.61, and 0.88 points. On FACED, removing the within-source relation raises macro-F1 but lowers AUROC from 87.77% to 75.29%, so its effect cannot be summarized by one metric. The full model has the highest mean in eight of the nine ablation columns; these are descriptive rankings, without a claim of statistical significance. The component-removal controls estimate contributions conditional on the remaining training objectives; combinations of removals would be needed to separate interactions among objectives.

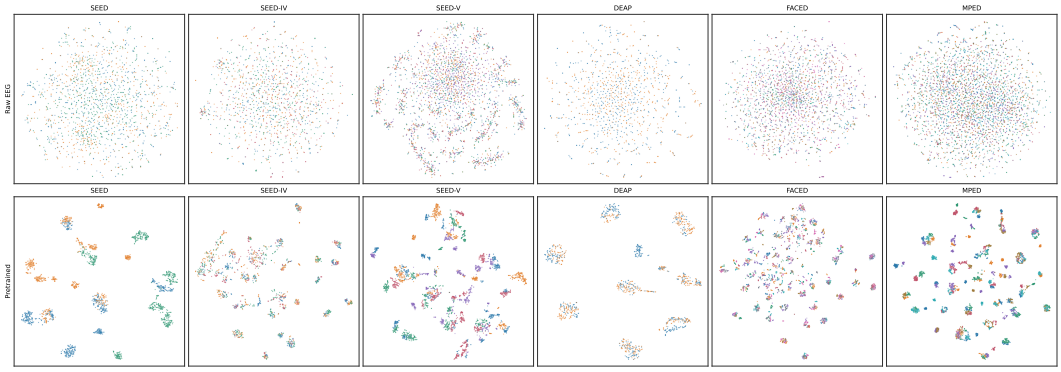

Figure 5: **Raw EEG and pretrained representations on six targets.** Columns show SEED, SEED-IV, SEED-V, DEAP, FACED, and MPED; rows show two-dimensional views of raw EEG and pretrained UniPER features. Colors denote each target's native classes.

Table 4: Ablation results on SEED-IV, DEAP, and FACED (mean ± standard deviation, %). Bold and underline mark the highest and second-highest means in each column. The zero-reference control removes both encoder and predictor reference subtractions; shuffled event pairs perturb the within-source relation.

Method	SEED-IV			DEAP			FACED		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Random initialization	45.97±0.92	45.02±5.26	55.65±1.82	58.93±0.52	59.35±4.15	70.63±1.78	27.38±0.60	28.92±0.55	50.23±1.79
Shuffled event pairs	66.11±1.87	61.88±6.53	70.44±1.84	80.11±1.74	80.35±4.15	87.51±2.37	56.91±1.71	57.15±0.68	81.41±1.66
w/o zero references	63.84±1.29	60.12±4.90	74.09±0.98	73.84±1.29	71.54±3.12	82.44±1.70	61.04±1.52	56.90±3.36	79.16±2.77
w/o local prediction	62.86±2.60	56.52±5.12	72.76±0.88	78.02±2.16	79.96±4.69	86.26±3.33	55.43±1.05	56.21±0.72	76.17±0.42
w/o context identity	66.32±2.19	67.15±3.69	78.61±0.81	79.44±2.64	80.27±1.96	89.09±3.36	58.66±1.48	57.21±2.25	70.52±3.51
w/o within-source relation	71.73±2.74	<u>72.18</u>±1.67	84.16±1.30	<u>85.39</u>±1.72	<u>84.55</u>±0.58	<u>91.77</u>±3.27	63.47±2.41	**66.82**±1.15	75.29±3.84
w/o cross-source relation	<u>71.75</u>±3.22	70.63±7.95	<u>85.59</u>±2.66	84.76±0.39	83.77±3.44	90.75±1.90	<u>64.72</u>±0.98	65.23±0.65	<u>86.48</u>±0.19
w/o SIGReg	69.14±1.67	69.51±5.45	83.79±1.65	80.96±0.66	81.65±4.24	84.17±1.91	58.24±0.81	58.26±0.88	81.78±1.60
UniPER	**74.34**±1.33	**74.09**±1.45	**88.08**±0.53	**87.37**±1.58	**87.17**±1.75	**92.97**±1.26	**65.60**±1.14	<u>65.82</u>±1.19	**87.77**±1.31

The random-initialization baseline scores 45.97%, 58.93%, and 27.38% BACC on SEED-IV, DEAP, and FACED. Source pretraining raises these values to 74.34%, 87.37%, and 65.60%, corresponding to gains of 28.37, 28.44, and 38.22 points. Shuffling event pairs gives 66.11%, 80.11%, and 56.91%, respectively. These comparisons examine the pretrained initialization and the recording-event structure used during learning.

Scope of the findings. The SD–SI gaps and the SEED-IV exception show that source pretraining leaves substantial variation across participants and targets. Moreover, the FACED analysis covers 51 of the dataset's 123 participants, and native label vocabularies differ across targets. The reported comparisons therefore characterize the evaluated partitions and readouts; they do not establish comparable performance on every participant population or in everyday affective monitoring.

4 Conclusion

UniPER couples position-referenced latent prediction with event-aware relation learning to produce a common EEG representation for affective tasks. It is designed for the case in which a target dataset changes participants, montage, stimuli, and emotion vocabulary at once. The results indicate that this source training helps subject-dependent adaptation across several targets, while the larger subject-independent gaps and the SEED-IV exception expose a remaining generalization limit. Such reuse could reduce labeled-data demands in affect-aware assistive systems, but responsible use requires privacy protection and assessment of uneven performance across participants. Future work should test prospective recordings from more diverse cohorts and develop calibration that remains reliable as sensors and contexts change.

AI USE DISCLOSURE

AI-assisted tools, including OpenAI Codex, were used to support language editing, literature search, code development and debugging, experiment scripting and monitoring, and figure preparation. The authors remain responsible for validating the code, experimental results, and references, and for all scientific claims, interpretations, and final content of this paper.

REFERENCES

Samannaya Adhikari, Nitin Choudhury, Swastika Bhattacharya, Nabamita Deb, Rajdeep Ghosh, Daisy Das, and Bhabesh Kalita. EEG Emotion Recognition Dataset_GU. Figshare record 29170289, 2025. Dataset, posted 29 May 2025.

Talha Burak Alakus, Murat Gonen, and Ibrahim Turkoglu. Database for an emotion recognition system based on EEG signals and various computer games—GAMEEMO. *Biomedical Signal Processing and Control*, 60:101951, 2020. doi: 10.1016/j.bspc.2020.101951.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023.

Alexei Baevski, Wei-Ning Hsu, Qiantong Xu, Arun Babu, Jiatao Gu, and Michael Auli. data2vec: A general framework for self-supervised learning in speech, vision and language. In *International Conference on Machine Learning*, pp. 1298–1312. PMLR, 2022.

Randall Balestriero and Yann LeCun. LeJEPA: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025. doi: 10.48550/arXiv.2511.08544.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing EEG dataset. *Scientific Data*, 10(1):740, 2023. doi: 10.1038/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.

Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. TSception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2023. doi: 10.1109/TAFFC.2022.3169001.

Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Su Xian Chong, Fang Ji, Nathanael Ren Jie Tong, Christopher Li Hsian Chen, and Juan Helen Zhou. Brain-JEPA: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. In *Advances in Neural Information Processing Systems*, volume 37, pp. 86048–86073, 2024. doi: 10.52202/079017-2732.

Erhan Ekmekcioglu and Yucel Cimtay. Loughborough university multimodal emotion dataset-2 (LUMED-2), 2021. URL `https://figshare.com/articles/dataset/Loughborough_University_Multimodal_Emotion_Dataset_-_2/12644033`. Dataset, version 5.

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.

Zitao Fang, Chenxuan Li, Hongting Zhou, Shuyang Yu, Guodong Du, Ashwaq Qasem, Yang Lu, Jing Li, Junsong Zhang, and Sim Kuan Goh. NeurIPT: Foundation model for neural interfaces. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL `https://openreview.net/forum?id=D4hcJPkJ3y`.

Jean-Bastien Grill, Florian Strub, Florent Altche, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284, 2020.

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.

Weibang Jiang, Liming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=QzTpTRVtrP.

Stamos Katsigiannis and Naeem Ramzan. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107, 2018. doi: 10.1109/JBHI.2017.2688239.

Sander Koelstra, Christian Muehl, 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.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15:653659, 2021. doi: 10.3389/fnhum.2021.653659.

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.

Rui Li, Yiting Wang, Wei-Long Zheng, and Bao-Liang Lu. A multi-view spectral-spatial-temporal masked autoencoder for decoding emotions with self-supervised learning. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 6–14, 2022. doi: 10.1145/3503161.3548243.

Yitong Li, Wenming Zheng, Yuan Zong, Zhen Cui, Tong Zhang, and Xiaoyan Zhou. Bidann: A bidirectional domain adaptation neural network for EEG-based emotion recognition. In *International Conference on Neural Information Processing*, pp. 132–141. Springer, 2018.

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.

Sudhakar Mishra, Md. Asif, Uma Shanker Tiwary, and Narayanan Srinivasan. DENS: Dataset on emotion with naturalistic stimuli, 2026. Dataset, version 1.0.0.

Navid Mohammadi Foumani, Geoffrey Mackellar, Soheila Ghane, Saad Irtza, Nam Nguyen, and Mahsa Salehi. EEG2Rep: Enhancing self-supervised EEG representation through informative masked inputs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5544–5555, 2024. doi: 10.1145/3637528.3671600.

Huy-Tung Phuong, Eun-Tack Im, Myeong-Seok Oh, and Gwang-Yong Gim. EEGEmotions-27: A large-scale EEG dataset annotated with 27 fine-grained emotion labels. *IEEE Access*, 13:176915–176932, 2025. doi: 10.1109/ACCESS.2025.3620677.

Xinke Shen, Xianggen Liu, Xin Hu, Dongrui 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.

Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. MPED: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7:12177–12191, 2019. doi: 10.1109/ACCESS.2019.2891579.

Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2020. doi: 10.1109/TAFFC.2018.2817622.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. doi: 10.1109/TNSRE.2022.3230250.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of EEG signals. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1239.

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 *International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=NPNUHgHF2w`.

Huqin Weng, Jiayang Huang, Yimin Wen, Jie Du, Chi-Man Vong, and Chuangquan Chen. Masked generative-contrastive representation learning for cross-dataset EEG-based emotion recognition. *arXiv preprint arXiv:2607.04139*, 2026. doi: 10.48550/arXiv.2607.04139.

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(1):1142, 2025. doi: 10.1038/s41597-025-05349-2.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal transformer for cross-data learning in the wild. In *Advances in Neural Information Processing Systems*, volume 36, pp. 78240–78260, 2023a. doi: 10.52202/075280-3420.

Chaoqi Yang, Cao Xiao, M. Brandon Westover, and Jimeng Sun. Self-supervised electroencephalogram representation learning for automatic sleep staging: Model development and evaluation study. *JMIR AI*, 2:e46769, 2023b. doi: 10.2196/46769.

Pei Yang, Niqi Liu, Xinge Liu, Yezhi Shu, Wenqi Ji, Ziqi Ren, Jenny Sheng, Minjing Yu, Ran Yi, Dan Zhang, and Yong-Jin Liu. A multimodal dataset for mixed emotion recognition. *Scientific Data*, 11:847, 2024. doi: 10.1038/s41597-024-03676-4.

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, pp. 53875–53891, 2023. doi: 10.52202/075280-2344.

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.

Li-Ming Zhao, Xu Yan, and Bao-Liang Lu. Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 863–870, 2021. doi: 10.1609/aaai.v35i1.16169.

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.

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.

Peixiang Zhong, Di Wang, and Chunyan Miao. EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2022. doi: 10.1109/TAFFC.2020.2994159.

Yijin Zhou, Fu Li, Yang Li, Youshuo Ji, Lijian Zhang, Yuanfang Chen, and Huaning Wang. Enhancing cross-dataset EEG emotion recognition: A novel approach with emotional EEG style transfer network. *IEEE Transactions on Affective Computing*, 16(3):2157–2171, 2025. doi: 10.1109/TAFFC.2025.3555439.

A Related Work

EEG pretraining and source domain. BENDR and CBraMod demonstrate reusable self-supervised representations from the clinical Temple University Hospital EEG corpus, using contrastive context prediction and masked signal reconstruction, respectively (Kostas et al., 2021; Wang et al., 2025). BIOT addresses heterogeneous biosignal formats; LaBraM, EEGPT, and REVE learn from broad EEG collections with masked modeling or representation alignment (Yang et al., 2023a; Jiang et al., 2024; Wang et al., 2024; El Ouahidi et al., 2025). Affective data also appear in broader pretraining settings: MMM jointly pretrains on an unlabeled SEED/SEED-IV union in its emotion experiments, and NeurIPT's mixed source corpus includes affective recordings (Yi et al., 2023; Fang et al., 2025). These models establish that EEG representations can cross recording domains. UniPER studies an affective multi-source setting with six held-out native-label emotion datasets, using masked latent prediction to learn a shared representation.

Emotion transfer and alignment. Task-trained EEGNet, EEG Conformer, TSception, DGCNN, and RGNN model temporal or spatial patterns for downstream decoding (Lawhern et al., 2018; Song et al., 2023; Ding et al., 2023; Song et al., 2020; Zhong et al., 2022). Domain adaptation methods such as BiDANN and plug-and-play adaptation address participant or domain shifts, while CLISA explicitly seeks subject-invariant emotion features by pairing responses to the same stimulus across participants (Li et al., 2018; Zhao et al., 2021; Shen et al., 2023). Cross-dataset transfer adds changes in acquisition and annotation; E2STN explicitly transfers emotion-related EEG style between datasets (Zhou et al., 2025). Existing multi-dataset affective pretraining is particularly close to our setting. MMM reconstructs a SEED/SEED-IV source union but evaluates on those dataset families; mdJPT uses label-free covariance and temporally aligned inter-subject objectives across emotional EEG; EMOD instead maps heterogeneous emotion labels into valence–arousal space for supervised contrastive learning (Yi et al., 2023; Zhang et al., 2025; Chen et al., 2026). These works use reconstruction, paired recording structure, or label semantics to organize transfer. UniPER uses source EEG without emotion labels, predicts masked teacher content as its main objective, and keeps all six target corpora outside source pretraining.

Predictive representations. Self-distilled and masked latent objectives in BYOL, data2vec, and I-JEPA provide precedents for learning from a slowly updated target rather than a hand-designed class label (Grill et al., 2020; Baevski et al., 2022; Assran et al., 2023). In affective EEG, MV-SSTMA reconstructs masked spectral, spatial, and temporal views (Li et al., 2022). EEG2Rep already predicts masked EEG latents, evaluates DREAMER emotion decoding, and tests cross-domain pretraining; EEGPT aligns masked spatiotemporal EEG representations (Mohammadi Foumani et al., 2024; Wang et al., 2024). MGCRL is an especially close affective predecessor: it combines JEPA-style masked latent prediction with contrastive learning, pretrains on FACED, and adapts to SEED-series datasets (Weng et al., 2026). Brain-JEPA extends predictive learning to fMRI dynamics (Dong et al., 2024). UniPER defines the EMA teacher's token content relative to a position-matched zero-waveform response and centers the predictor with a zero-memory reference under identical queries. The corresponding complete-input online content is pooled into the transfer representation.

Regularizing the exported representation. UniPER connects hidden-token learning to the exported window vector by applying characteristic-function regularization, SIGReg, to source-centered online window embeddings, following the distributional criterion developed for LeJEPA (Balestriero & LeCun, 2025). Context identity and within-source ranking compare aggregated hidden predictions with hidden teacher targets. Cross-source relations use these predictive summaries as queries and complete-window endpoints as keys. The resulting objectives connect local content prediction, recording-event structure, and the window representation within one multi-source learning framework.

B Structured Masking

Each preprocessed window contains 25 non-overlapping temporal patches per observed electrode. Let V denote the valid sensor–time positions. The channel mask selects a cluster of neighboring electrodes at an approximately 28% rate; the temporal mask selects a contiguous interval at an approximately 28% rate. Their union gives the hidden positions $M = (M_C \cup M_T) \cap V$, and

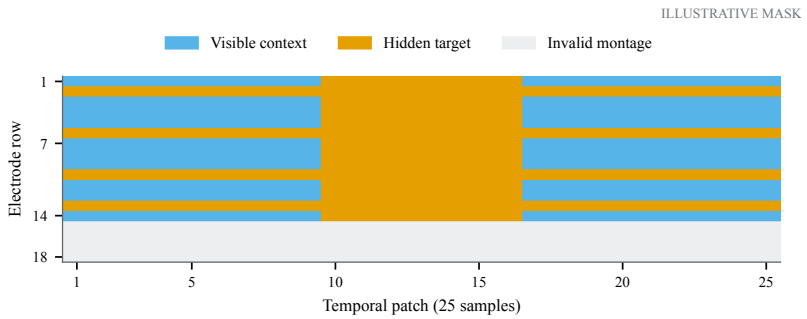

Figure B.1: **Illustrative sensor–time mask.** Blue cells are visible context, orange cells are hidden prediction targets, and gray cells are invalid montage positions. Rows represent electrodes and columns represent 25-sample temporal patches.

$H = V \setminus M$ supplies visible context. The recorded hidden fraction is approximately 47%. Spatial proximity is defined through electrode geometry. The encoder processes the measured montage directly.

The patch stem applies its convolutions within individual 25-sample patches. Structured masking defines the online context for latent prediction, while the teacher encodes the complete preprocessed window. The masked online path, complete online path, and complete teacher path each subtract the response to a zero waveform under matching coordinates and validity conditions. The predictor uses the same hidden-position queries with real and zero memory, subtracts the two outputs, and applies LayerNorm followed by L2 normalization. It processes at most 128 hidden queries at a time. Figure B.1 illustrates the sensor–time geometry.

C Training Configuration

Table C.1 specifies the encoder, predictor, and objectives; Table C.2 gives pretraining and SD adaptation settings. Source pretraining used two NVIDIA H100 GPUs on HPC3, with one source batch of 12 windows per GPU at each of 7,200 optimizer updates. This yields 14,400 source batches and 1,600 batches per source. Pairable sources use three participants and four events per batch; the remaining sources use four participants and three windows per participant. Target adaptation and evaluation used one H100 GPU per run.

The online encoder has 6,465,376 parameters. The total trainable pretraining parameter count is 7,257,696; including the EMA teacher gives 13,723,072 parameters. AdamW uses a fixed source learning rate, and the teacher is updated after each optimizer step. Sixteen content tokens per window are sampled for representation-health diagnostics; the optimized objectives are those in Eq. 10.

SD adaptation initializes a new $512 \to K$ linear head for a target with K classes and updates it jointly with the online encoder. Encoder and head parameter groups use separate learning rates. Training runs for six epochs and evaluates the final epoch. The three partition seeds are `2026091601`, `2026091602`, and `2026091603`. Appendix E.1 describes the SD, SI, and zero-shot evaluation procedures.

Table C.1: Encoder, predictor, masking, and objective settings. Widths denote token feature dimensions. All relation temperatures are 0.20.

Component	Setting
Window / rate / patch	5 s / 125 Hz / 25 samples; 25 patches per channel
Reference / bandpass	Measured-channel average / fourth-order 0.5–45 Hz Butterworth, zero phase
Input scaling	Per-channel demeaning; whole-window RMS scaling
Montage	3D electrode coordinates and valid-channel mask
Patch stem	Patch-local kernels 3, 7, 15; eight filters each; linear projection to 192D
Position encoding	Linear electrode-coordinate embedding + sinusoidal time embedding
Context backbone	Three spatial-then-temporal blocks; 192D; six heads
Endpoint	$192 \rightarrow 512$ linear + GELU; one 512D, eight-head spatiotemporal block; $512 \rightarrow 512$ readout
FFN expansion / dropout	$2\times$ / no dropout configured
Exported representation	Valid complete-input content-token mean + LayerNorm; 512D
Structured mask	Neighboring-electrode cluster 28%; continuous time interval 28%; union; observed hidden fraction $\approx 47\%$
Prediction memory	Visible 512D content tokens; $512 \rightarrow 192$ projection
Prediction blocks / output	Two 192D, six-head cross-attention blocks; $192 \rightarrow 512$ output
Hidden queries	At most 128 processed at a time
Encoder reference	Matched zero-waveform subtraction in masked online, full online, and full teacher views
Predictor reference	Same queries and zero memory; subtract reference output
Prediction / target normalization	LayerNorm then L2 normalization
Teacher	Online initialization; stopped gradients; fixed EMA momentum 0.996
Masked JEPA	Mean position-wise cosine distance
Context identity	Source-local multi-positive prediction-to-teacher event matching, including self-pairs
Within-source relation	Cross-participant same-event positive; a different stimulus from the positive participant as negative; softplus cosine ranking
Cross-source relation	Predictive queries to other-source endpoint keys; teacher-to-student KL; ordered source pairs weighted equally
Endpoint SIGReg	Source-centered full online window vectors; 256 directions; 17 integration points on $[0, 3]$
JEPA / identity / SIGReg weights	1 / 1 / 1
Relation coefficient $r(t)$	0 through 20%; linear increase to 0.05 at 40%; then 0.05
Diagnostic token sample	16 content tokens per window

Table C.2: Pretraining and target adaptation settings. SD partitions complete trials within participants; SI partitions participants.

Hyperparameter	Value
Source pretraining	
Optimizer / learning rate	AdamW / fixed 2×10^{-4}
Weight decay / updates	0.05 / 7,200
Hardware	HPC3; two NVIDIA H100 GPUs
Source batch / population	12 windows / 24 windows per optimizer update
Pairable-source batch	Three participants $\times$ four events
Other-source batch	Four participants $\times$ three windows
Source balance	1,600 batches per source across nine sources
Checkpoint seed	19260917
Precision / gradient clipping	CUDA BF16 autocast / global norm 5
EMA update	After the optimizer step; momentum 0.996
Subject-dependent target adaptation	
Hardware	HPC3; one NVIDIA H100 GPU per run
Trial grouping	Complete (subject, session, trial) groups within each participant
Training / test fraction	Approximately 75% / 25% of trials
Split seeds	2026091601 / 2026091602 / 2026091603
Classifier	Fresh $512 \rightarrow K$ linear head
Encoder / head learning rates	2×10^{-5} / 5×10^{-4}
Optimizer / weight decay	AdamW / 10^{-4}
Batch size / epochs	16 / 6
Checkpoint selection	Final epoch
Precision / cache-loader workers	BF16 / 0
Subject-independent target adaptation	
Training / test fraction	Approximately 75% / 25% of participants; three repetitions
Encoder / head learning rates	2×10^{-5} / 5×10^{-4}
Training budget	6 epochs

D BASELINES AND SUPERVISION

The supervised group contains EEGNet, EEG Conformer, TSception, DGCNN, and RGNN (Lawhern et al., 2018; Song et al., 2023; Ding et al., 2023; Song et al., 2020; Zhong et al., 2022). CLISA, ContraWR, and MMM form the self-supervised group (Shen et al., 2023; Yang et al., 2023b; Yi et al., 2023). The pretrained group contains BENDR, BIOT, LaBraM, CBraMod, EMOD, and mdJPT (Kostas et al., 2021; Yang et al., 2023a; Jiang et al., 2024; Wang et al., 2025; Chen et al., 2026; Zhang et al., 2025). These categories describe each method's training approach and role in the comparison. EMOD uses affective label guidance during its original pretraining. UniPER uses recording-event metadata during source pretraining and target labels during downstream evaluation.

E DATASET ALLOCATION AND SPLITS

The nine pretraining sources are DREAMER (Katsigiannis & Ramzan, 2018), EmoEEG-MC (Xu et al., 2025), MDME (Yang et al., 2024), DENS (Mishra et al., 2026), GAMEEMO (Alakus et al., 2020), EEGEmotions-27 (Phuong et al., 2025), GU (Adhikari et al., 2025), LUMED-2 (Ekmekcioglu & Cimtay, 2021), and SEED-VII (Jiang et al., 2025). GU denotes the Gauhati University EEG Emotion Recognition Dataset. The retained source corpus contains 77,850 five-second windows. Recording-event keys use session, stimulus, and window identifiers within each source.

The held-out targets are SEED and SEED-IV (Zheng & Lu, 2015; Zheng et al., 2019), SEED-V (Liu et al., 2022), DEAP (Koelstra et al., 2012), FACED (Chen et al., 2023), and MPED (Song et al., 2019). Their caches contain 97,985 windows in total. Table 1 reports the retained cache sizes alongside original dataset participant counts. The FACED cache contains 8,568 windows from 51 participants, `sub000`–`sub050`, out of the original 123. The analyzed FACED recordings are at 250 Hz; EmoEEG-MC includes recordings at both 600 and 1200 Hz.

Common signal processing. Each dataset is read at the sampling rate of its available recordings. We apply measured-electrode average referencing, a per-window fourth-order 0.5–45 Hz zero-phase Butterworth bandpass, resampling to 125 Hz, per-channel demeaning, and whole-window RMS scaling, in this order. Non-finite windows are rejected. Every retained window has 625 samples per channel. Dataset-provided preprocessing remains part of each input's acquisition history. Three-dimensional electrode coordinates and an explicit validity mask describe the measured montage.

E.1 TARGET EVALUATION PROTOCOLS

All six target datasets are excluded from source pretraining. Each target retains its native label vocabulary. Target readouts use the complete-input online representation in Eq. 2; the masked predictor and EMA teacher are used during source pretraining.

Subject-dependent (SD). Partitions operate within each participant and preserve complete $(\text{subject}, \text{session}, \text{trial})$ groups. Approximately 75% of trials form the training partition and 25% the test partition; all windows from a trial share its assignment, so windows from the same trial never cross the boundary. The three fixed partition seeds are `2026091601`, `2026091602`, and `2026091603`. For each partition, adaptation starts from the source-pretrained online encoder and a fresh $512 \to K$ linear head, where K is the target's class count. AdamW jointly updates the encoder and head for six epochs with batch size 16, weight decay 10^{-4}, and respective learning rates 2×10^{-5} and 5×10^{-4}. Evaluation uses the final epoch, rather than selecting an epoch by test performance (Table C.2).

Subject-independent (SI). SI uses three separate participant-disjoint target partitions, each assigning approximately 75% of participants to adaptation and 25% to test. The partition unit is the participant: all sessions, trials, and windows of each held-out participant remain on the test side, and no participant occurs in both sets. In each repetition, the target encoder and classifier are adapted on labeled training participants using the same encoder and classifier learning rates and six-epoch budget as SD, then evaluated on held-out participants. SI partitions are distinct from the within-participant SD partitions.

Zero-shot gallery evaluation. The source-pretrained online encoder is frozen and maps both reference and query windows to the same 512D representation in Eq. 2. A labeled gallery from the target dataset supplies the class vocabulary and reference examples for classifying target queries; query labels are used only to compute evaluation metrics. Neither the encoder nor a target classifier is updated in this protocol. Here zero-shot denotes transfer without target parameter updates, while target labels enter through the gallery.

F Full Target Results

Tables F.1–F.6 report SD, SI, and zero-shot results for each target. SD separates trials within participants; SI separates participants. The SD and zero-shot columns correspond to Tables 2 and 3, respectively. A dash denotes an unreported comparison.

Table F.1: SEED results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	71.09±1.42	70.70±1.62	79.74±1.06	43.04±1.62	42.11±1.50	59.80±2.50	—	—	—
EEG Conformer	57.30±4.34	56.57±5.75	72.69±2.22	44.06±3.43	42.63±3.75	61.74±2.70	—	—	—
TSception	53.02±1.73	52.72±1.96	68.19±1.74	44.56±1.21	43.57±1.38	61.15±1.44	—	—	—
DGCNN	76.68±0.82	76.05±0.57	82.30±0.47	39.09±1.72	35.25±2.77	57.02±1.24	—	—	—
RGNN	40.15±1.67	37.56±4.16	55.27±1.46	36.78±3.61	32.27±7.29	53.41±1.30	—	—	—
SSL									
CLISA	78.76±1.01	78.13±0.99	84.46±0.68	63.08±1.61	62.17±1.46	72.81±1.49	—	—	—
ContraWR	78.19±1.05	77.66±1.03	84.27±0.72	61.73±1.70	61.40±1.50	73.29±1.43	—	—	—
MMM	79.31±0.72	78.69±0.87	84.84±0.48	64.99±2.10	64.13±1.94	73.87±1.78	—	—	—
Pretraining									
BENDR	76.23±1.23	75.91±1.02	82.93±0.84	57.93±1.84	57.57±1.97	71.00±1.80	37.93±1.49	36.90±1.34	54.75±1.00
BIOT	78.52±0.68	77.86±0.86	83.45±0.39	61.20±2.27	60.30±2.11	71.41±1.77	41.57±1.21	40.45±1.43	57.84±0.79
LaBraM	80.25±1.34	79.39±1.15	84.60±0.89	63.69±2.11	63.34±2.01	74.44±2.01	42.24±1.62	41.08±1.70	58.34±1.45
CBraMod	79.35±0.66	78.71±0.54	84.50±0.39	66.53±2.21	65.61±2.10	74.71±2.14	47.12±1.59	46.47±1.59	61.99±1.36
EMOD	80.80±0.83	80.36±0.73	84.99±0.49	66.98±2.27	66.41±2.14	76.49±1.81	51.01±1.41	50.18±1.21	65.21±1.14
mdJPT	80.85±1.23	80.43±1.42	84.34±0.73	68.05±2.25	67.66±2.21	76.94±1.99	56.29±1.35	55.21±1.19	68.79±0.87
UniPER	**84.33**±1.96	**84.23**±1.51	**87.95**±1.30	**70.03**±1.98	**69.56**±2.03	**79.20**±1.69	**58.29**±1.28	**57.19**±1.47	**75.54**±1.61

Table F.2: SEED-IV results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	59.67±1.79	58.68±1.56	80.00±1.52	38.57±2.91	38.03±3.04	65.96±2.66	—	—	—
EEG Conformer	63.37±1.10	62.26±1.05	82.88±0.84	43.19±2.65	42.70±2.59	68.25±2.12	—	—	—
TSception	65.79±1.33	65.33±1.51	84.46±0.80	45.62±2.92	45.08±2.80	69.90±2.42	—	—	—
DGCNN	68.44±1.76	67.64±1.79	85.84±1.30	48.36±2.64	47.80±2.66	72.44±2.55	—	—	—
RGNN	69.63±1.54	68.96±1.39	86.75±1.41	50.11±2.43	49.18±2.60	74.27±2.02	—	—	—
SSL									
CLISA	72.04±1.91	71.09±2.00	88.28±1.63	52.86±2.96	52.35±3.04	76.24±2.57	—	—	—
ContraWR	70.21±1.34	69.42±1.34	87.00±1.12	51.42±2.72	50.70±2.45	74.68±2.27	—	—	—
MMM	72.48±1.90	71.59±2.01	88.49±1.68	55.26±2.56	54.20±2.86	76.22±2.37	—	—	—
Pretraining									
BENDR	68.16±1.20	67.09±1.19	85.47±1.00	47.27±2.79	46.16±2.99	72.77±2.35	31.22±1.75	29.69±1.97	57.90±1.57
BIOT	70.60±1.75	69.44±1.60	86.16±1.45	50.47±2.84	49.52±2.62	73.26±2.78	34.72±2.10	33.24±1.92	60.88±1.78
LaBraM	73.06±1.18	72.00±1.16	87.68±0.76	56.50±2.73	55.54±2.51	77.14±2.36	40.26±1.37	39.04±1.67	65.62±0.84
CBraMod	74.73±1.87	73.56±1.65	**89.21**±1.65	53.56±2.41	52.91±2.45	75.85±2.02	38.11±1.49	36.89±1.61	63.69±1.06
EMOD	**75.55**±1.85	**74.79**±1.78	88.80±1.66	58.09±2.25	57.36±2.27	78.09±2.07	43.17±1.66	41.89±1.62	67.93±1.32
mdJPT	74.96±1.06	73.73±1.29	89.02±0.77	**58.30**±2.58	**57.60**±2.86	**78.52**±2.50	**46.10**±2.24	**44.52**±2.42	**70.08**±2.05
UniPER	74.34±1.33	74.09±1.45	88.08±0.53	57.92±2.26	57.11±2.91	78.24±2.29	45.28±1.40	44.15±1.74	68.57±1.83

Table F.3: SEED-V results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	50.06±1.26	49.09±1.09	72.70±1.15	31.76±2.21	30.66±1.97	59.90±1.70	—	—	—
EEG Conformer	55.20±1.36	54.31±1.15	74.54±1.03	35.68±2.71	34.66±2.94	62.88±2.48	—	—	—
TSception	56.55±1.55	55.98±1.63	75.74±1.41	38.01±2.03	37.39±1.88	64.31±1.82	—	—	—
DGCNN	59.75±1.23	58.94±1.10	76.99±0.95	41.50±2.79	40.53±2.93	67.00±2.58	—	—	—
RGNN	60.00±1.54	58.88±1.74	77.59±1.11	42.44±2.09	41.58±2.34	66.87±2.03	—	—	—
SSL									
CLISA	62.89±1.45	62.29±1.59	78.89±0.97	46.22±2.57	45.38±2.49	69.12±2.09	—	—	—
ContraWR	61.62±1.43	60.82±1.54	78.48±1.24	44.86±2.44	44.41±2.44	68.83±2.12	—	—	—
MMM	63.87±1.12	63.12±1.05	79.85±1.06	52.54±2.30	51.62±2.38	72.57±2.22	—	—	—
Pretraining									
BENDR	58.97±1.59	57.89±1.86	76.45±1.18	39.72±2.78	38.87±2.86	65.08±2.65	26.31±1.86	24.50±1.84	55.64±1.74
BIOT	62.18±1.13	61.35±1.20	79.43±1.01	43.40±2.58	42.99±2.49	67.54±2.08	30.30±1.81	28.84±2.00	57.49±1.72
LaBraM	64.35±1.79	63.60±1.64	79.64±1.24	49.73±2.38	49.28±2.64	71.68±2.31	32.16±1.45	30.79±1.50	59.22±1.36
CBraMod	65.84±1.52	64.85±1.59	80.33±1.20	54.24±2.05	53.61±2.21	74.20±1.88	34.73±1.90	33.85±2.00	60.65±1.49
EMOD	65.84±1.62	65.65±1.59	81.14±1.57	55.49±2.00	54.79±2.07	73.57±1.70	37.88±1.98	36.58±2.20	62.27±1.76
mdJPT	65.96±1.46	65.80±1.68	81.50±0.95	57.00±2.66	56.35±2.61	75.52±2.37	42.60±1.32	41.74±1.19	65.57±1.19
UniPER	**71.46**±1.17	**70.21**±1.46	**84.01**±0.99	**60.27**±2.38	**59.97**±2.89	**78.79**±1.39	**45.87**±1.35	**44.87**±1.31	**66.39**±1.57

Table F.4: DEAP results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	72.41±1.55	71.75±1.41	80.31±1.07	49.27±1.55	46.25±2.15	49.31±2.00	—	—	—
EEG Conformer	75.13±2.00	74.64±1.97	81.96±1.81	49.93±0.30	37.45±7.41	49.61±1.12	—	—	—
TSception	52.83±0.65	52.70±0.62	53.89±0.56	62.14±2.38	61.91±2.65	68.17±2.28	—	—	—
DGCNN	77.71±1.49	77.35±1.50	83.91±0.95	62.74±2.86	62.57±2.87	68.79±2.77	—	—	—
RGNN	78.57±1.27	78.14±1.44	86.03±0.86	64.18±2.32	64.02±2.41	71.41±1.84	—	—	—
SSL									
CLISA	80.41±1.41	80.23±1.63	86.53±0.86	66.51±2.34	65.72±2.56	74.26±1.96	—	—	—
ContraWR	79.43±1.12	78.98±1.15	86.70±0.60	65.47±2.83	64.75±2.89	72.19±2.61	—	—	—
MMM	82.19±1.78	82.08±1.73	89.16±1.32	71.86±2.38	71.22±2.12	78.24±2.03	—	—	—
Pretraining									
BENDR	79.09±1.66	78.88±1.80	85.02±1.26	64.45±2.67	64.11±2.67	72.14±2.51	55.35±2.26	54.70±2.31	59.30±1.83
BIOT	80.97±1.88	80.47±2.12	87.62±1.31	66.56±2.79	65.80±2.83	73.81±2.64	57.76±1.42	56.57±1.38	63.68±1.25
LaBraM	82.74±1.12	81.94±0.96	89.05±1.00	71.36±2.85	72.78±3.10	79.03±2.26	62.70±1.49	61.64±1.40	69.73±1.21
CBraMod	83.79±1.15	83.00±1.40	89.13±1.10	70.53±2.27	69.71±2.07	76.48±1.71	64.31±1.97	63.33±1.72	70.54±1.78
EMOD	83.94±1.45	83.40±1.74	89.03±1.30	74.58±2.60	74.00±2.71	80.19±2.45	66.69±2.22	66.11±2.43	73.24±1.98
mdJPT	84.92±1.43	84.20±1.46	90.52±0.95	75.09±3.05	74.65±2.82	81.31±2.92	68.28±1.59	67.79±1.63	**75.13**±1.32
UniPER	**87.37**±1.58	**87.17**±1.75	**92.97**±1.26	**77.54**±2.01	**77.24**±2.63	**83.76**±2.27	**70.73**±1.59	**70.68**±1.68	71.01±1.25

Table F.5: FACED results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	43.55±1.23	42.46±1.02	72.53±1.00	17.81±2.17	15.62±1.95	56.34±1.87	—	—	—
EEG Conformer	11.95±1.70	3.47±1.62	58.97±1.68	11.53±1.28	3.64±1.24	57.08±1.17	—	—	—
TSception	27.77±0.65	27.07±0.95	67.31±0.51	22.19±0.49	22.07±0.53	62.11±0.51	—	—	—
DGCNN	10.53±0.07	2.63±0.31	51.66±1.57	10.57±0.04	2.75±0.13	50.63±0.31	—	—	—
RGNN	52.06±1.63	50.39±1.83	77.76±1.16	10.55±0.00	2.64±0.00	49.69±0.35	—	—	—
SSL									
CLISA	54.02±1.48	52.53±1.60	78.46±1.09	29.96±2.75	28.12±2.66	64.43±2.57	—	—	—
ContraWR	53.43±1.86	51.83±1.64	79.06±1.78	28.45±2.84	26.88±2.88	64.17±2.57	—	—	—
MMM	55.39±1.25	53.92±1.02	78.86±1.11	32.52±2.29	30.42±2.21	65.77±2.20	—	—	—
Pretraining									
BENDR	50.86±1.15	49.43±0.96	76.50±0.98	25.74±2.83	23.74±2.84	62.24±2.28	12.80±1.55	10.43±1.31	51.37±1.32
BIOT	53.81±1.37	52.38±1.46	78.72±1.14	29.02±2.28	26.94±2.15	65.25±1.72	13.64±2.21	11.42±2.11	51.92±2.02
LaBraM	58.08±1.81	56.66±1.84	80.53±1.68	31.40±2.39	29.08±2.64	65.51±2.00	16.79±1.58	12.21±1.35	53.61±1.05
CBraMod	57.32±1.46	54.63±1.62	79.13±1.20	32.54±2.57	31.45±2.79	68.23±2.46	17.36±1.91	14.39±1.67	54.33±1.66
EMOD	59.63±1.90	58.52±1.81	80.95±1.44	39.04±2.75	37.64±2.88	71.01±2.68	18.72±1.85	16.66±1.68	57.00±1.68
mdJPT	58.43±1.38	56.90±1.31	80.09±1.09	37.30±2.37	35.65±2.42	68.47±2.15	15.39±1.90	12.88±1.75	54.04±1.68
UniPER	**65.60**±1.14	**65.82**±1.19	**87.77**±1.31	**44.06**±2.31	**44.71**±2.25	**75.63**±1.28	**22.85**±1.48	**22.26**±1.24	**63.75**±1.46

Table F.6: MPED results by subject-dependent, subject-independent, and zero-shot protocol. Cells are mean ± standard deviation percentages for BACC, macro-F1, and macro AUROC; dashes denote unreported zero-shot comparisons.

Method	Subject-dependent			Subject-independent			Zero-shot		
	BACC	F1	AUC	BACC	F1	AUC	BACC	F1	AUC
Supervised Methods									
EEGNet	46.83±1.14	45.55±1.02	73.99±0.86	23.42±2.53	21.86±2.54	58.81±1.96	—	—	—
EEG Conformer	51.28±1.65	49.99±1.43	76.50±1.08	27.89±2.86	26.22±2.78	62.18±2.78	—	—	—
TSception	53.95±1.82	52.70±2.00	77.05±1.67	30.57±2.71	29.14±2.76	64.51±2.31	—	—	—
DGCNN	56.59±1.30	55.51±1.45	78.57±0.85	16.06±1.36	12.34±1.18	50.99±0.92	—	—	—
RGNN	57.30±1.83	56.36±1.75	79.83±1.76	14.37±0.92	10.52±0.92	48.36±0.78	—	—	—
SSL									
CLISA	59.99±1.26	58.98±1.19	81.09±0.95	38.98±2.36	37.94±2.56	69.49±1.86	—	—	—
ContraWR	58.85±1.57	57.80±1.50	80.91±1.02	37.44±2.32	36.05±2.19	67.41±1.99	—	—	—
MMM	61.66±1.81	60.42±1.73	80.89±1.58	42.79±2.83	41.44±2.58	71.23±2.68	—	—	—
Pretraining									
BENDR	56.96±1.80	56.13±1.72	79.88±1.63	33.83±2.68	32.68±2.96	65.30±2.60	19.80±1.83	17.84±1.76	55.21±1.30
BIOT	59.81±1.47	58.77±1.30	80.36±1.11	36.75±2.57	35.65±2.67	68.48±2.38	22.80±2.19	20.65±1.94	57.10±1.72
LaBraM	61.79±1.83	60.71±2.05	81.08±1.36	41.71±2.47	40.76±2.70	70.46±2.36	24.15±1.72	21.74±1.47	58.60±1.31
CBraMod	63.49±1.70	62.50±1.69	82.20±1.29	44.93±2.77	43.44±2.74	73.25±2.66	27.31±2.16	25.62±2.12	60.38±1.82
EMOD	64.77±1.69	63.42±1.51	83.26±1.24	47.46±2.58	45.89±2.33	73.20±2.34	30.15±1.76	27.87±1.48	61.93±1.31
mdJPT	64.02±1.64	62.66±1.90	83.36±1.51	46.19±2.46	44.88±2.27	72.90±2.27	28.82±1.82	26.79±1.84	62.06±1.46
UniPER	**65.22**±1.23	**65.74**±1.24	**85.34**±1.29	**47.82**±2.27	**47.51**±2.69	**74.82**±2.30	**30.46**±1.24	**30.57**±1.77	**67.16**±1.15