CAN CROSS-IDENTITY SESSION-SHIFT TRANSFER IMPROVE EEG IDENTIFICATION ACROSS SESSIONS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-session EEG identification requires identity-discriminative representations to remain stable despite changes in cognitive state and acquisition conditions. We propose SHIFT-ID, a cross-identity session-shift transfer framework for the setting in which registered users provide no calibration data in their later evaluation sessions. SHIFT-ID constructs an empirical drift bank from paired sessions of auxiliary identities in the space of feature means and log standard deviations. It then transfers drifts across identities to generate session-informed training views and combines cosine consistency between representations with worst-view prototype regularization to preserve identity structure under these perturbations. At inference time, the fixed encoder, projection head, and classifier process each query independently, without calibration recordings from evaluation users or online parameter updates. We establish a unified information-isolation protocol on MCD, M3CV, and SEED-IV, evaluate twelve commonly used or recent EEG architectures, and perform information-budget-matched comparisons, transfer-direction analyses, mechanism ablations, and perturbation diagnostics. All experiments are repeated five times, and performance is reported as the mean over the five runs. The full configuration achieves mean Rank-1 accuracies of 70.94%, 77.83%, and 85.06% on the three datasets, respectively. Relative to CAGCNet, the strongest source-only external baseline under our protocol, the corresponding gains are 0.55, 0.42, and 0.38 percentage points. Under the same backbone and auxiliary-target budget, the Rank-1 gains over DANN are 0.63, 0.31, and 0.27 percentage points. Across the evaluated conditions, the directional, ablation, and perturbation analyses consistently support the contribution of empirical drift transfer and robust prototype regularization, while the modest absolute margins motivate a conservative interpretation of the improvements.

1 INTRODUCTION

EEG contains individual differences that can support biometric identification (Campisi & La Rocca, 2014; La Rocca et al., 2014), but enrollment and subsequent use rarely occur under identical conditions. Cognitive state, recording time, and electrode contact may change. Source-session classification can consequently exploit acquisition-specific correlations alongside identity information, making independent-session evaluation essential (Maiorana et al., 2016).

Transfer learning uses shared structure across subjects or sessions to reduce calibration burden (Jayaram et al., 2016); Riemannian approaches, for example, address distribution changes in EEG covariance matrices (Zanini et al., 2018). We consider a different information constraint: later recordings of the evaluation users are sealed, but paired historical sessions from other, auxiliary users are available. Can variation observed in those auxiliary users improve recognition of the enrolled evaluation identities without their target-session calibration data?

Two difficulties motivate our approach. Aligning entire session distributions may suppress identity-discriminative differences, while unconstrained random augmentation need not resemble observed session variation. Moreover, session effects can interact with identity, so an estimated shift cannot be assumed to be a pure session factor. We instead estimate each auxiliary user's within-identity difference in feature-statistic space and transfer the resulting empirical perturbation to other identities.

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.

Cosine consistency preserves representation orientation across the generated views; worst-view prototype classification encourages discrimination for the hardest sampled view.

Our setting is closed-set identification over all enrolled identities. *Recalibration-free* means that evaluation-user target data are excluded from training, normalization fitting, model selection, and test-time updating. Auxiliary target recordings remain available, so this is not source-only domain generalization. The drift bank and prototypes are training components; inference retains only the fixed encoder, projection head, and classifier.

The contributions are: (i) an information-isolation protocol separating auxiliary histories from evaluation target sessions; (ii) paired cross-identity drift transfer with consistency and worst-view prototype regularization; and (iii) comparisons across three datasets, twelve external architectures, matched-information strategies, transfer directions, and controlled diagnostics. We distinguish the effect of additional training information from the incremental effects of the proposed constraints and interpret small absolute improvements conservatively.

2 Related Work

EEG representations. General EEG decoding spans convolutional and attention-based models (Lotte et al., 2018; Roy et al., 2019). Our comparisons include DeepConvNet and ShallowConvNet (Schirrmeister et al., 2017), EEGNet (Lawhern et al., 2018), EEG-TCNet (Ingolfsson et al., 2020), FBCNet (Mane et al., 2021), TSception (Ding et al., 2023), ATCNet (Altaheri et al., 2023), EEG Conformer (Song et al., 2023), EEG-SimpleConv (El Ouahidi et al., 2023), EEG-Deformer (Ding et al., 2025), and TCFormer (Altaheri et al., 2025). Models designed for motor imagery or emotion recognition must be reevaluated for identity recognition. Identity-oriented work includes adversarial cross-session biometrics (Özdenizci et al., 2019) and CAGCNet (Wang et al., 2025); the latter is included as a recent external baseline.

Statistical augmentation and identity constraints. AdaIN modifies feature moments (Huang & Belongie, 2017); MixStyle mixes statistics (Zhou et al., 2021), and DSU models their uncertainty (Li et al., 2022). Our perturbations instead use paired, within-identity session differences and preserve correspondence among statistical coordinates. Unlike mixup (Zhang et al., 2018), the current sample's identity label is unchanged. Our consistency term operates on normalized representations, whereas AugMix uses prediction consistency (Hendrycks et al., 2020). Prototypical networks (Snell et al., 2017) and supervised contrastive learning (Khosla et al., 2020) motivate class-structured representations, but our prototypes are batch regularizers built from original and shifted views, not independent test-time enrollment templates. Appendix C gives further context.

3 Method

Figure 1 provides an overview of the proposed SHIFT-ID framework. The framework consists of four stages: the data and evaluation protocol, paired session-drift estimation from auxiliary identities, cross-identity drift transfer with discriminative regularization, and fixed inference without evaluation-user calibration or test-time adaptation.

3.1 Problem Definition and Statistical Feature Encoding

Let $\mathcal{Y}$ denote the set of enrolled identities, partitioned into disjoint auxiliary and evaluation sets: $\mathcal{Y} = \mathcal{A} \cup \mathcal{E}$ and $\mathcal{A} \cap \mathcal{E} = \varnothing$. Source-session data $\mathcal{S}$ contain all identities. Only the auxiliary target-session data $\mathcal{T}_{\mathcal{A}}$ are available during learning; the evaluation target-session data $\mathcal{T}_{\mathcal{E}}$ are reserved for testing and excluded from training, normalization fitting, and model selection. Each test sample is classified over the full $|\mathcal{Y}|$-class candidate space, so assigning an evaluation sample to an auxiliary identity also counts as an error.

For an input $x \in \mathbb{R}^{C \times T}$, channel-wise normalization uses fixed means and standard deviations estimated from the source training set. The encoder F_θ applies a 1×63 temporal convolution, a spatial convolution spanning all C input channels, and a depthwise temporal convolution of length 15 followed by a pointwise convolution. The channel widths are 16, 32, and 32; each block uses

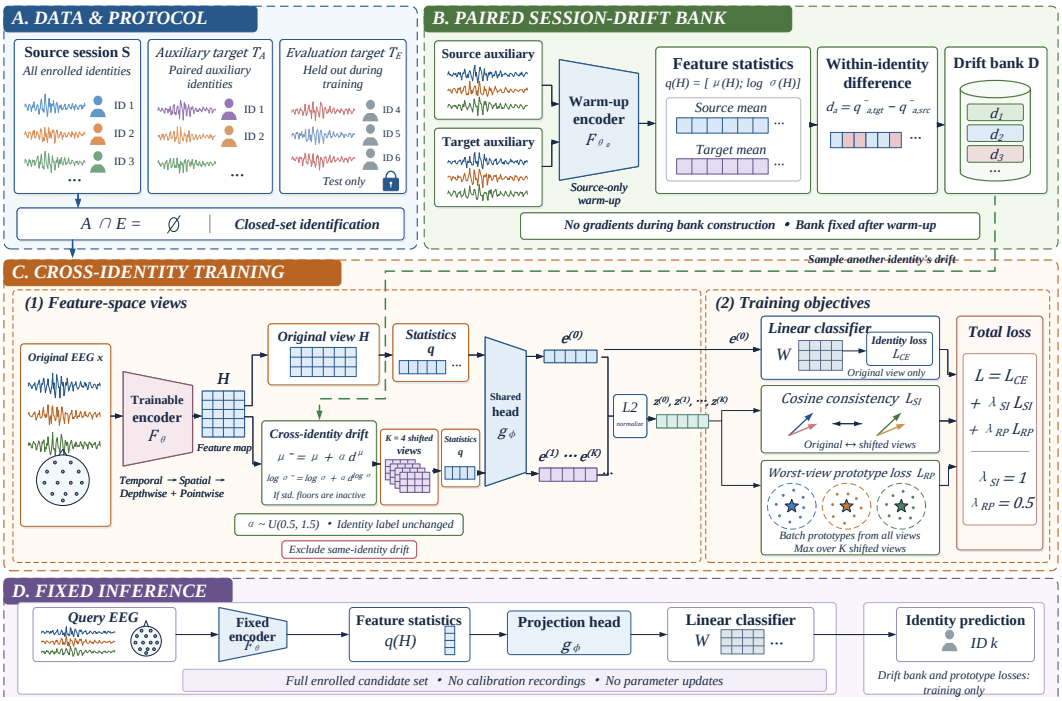


Figure 1: Overview of SHIFT-ID. **(A)** Source-session data from all enrolled identities are used for supervised training, whereas paired target-session data are available only for auxiliary identities and evaluation target sessions remain strictly held out. **(B)** A fixed drift bank is constructed from within-identity source-to-target differences in feature mean and log-standard-deviation statistics. **(C)** Cross-identity drifts generate session-shifted feature views, which are jointly optimized using identity cross-entropy, cosine consistency, and worst-view prototype regularization. **(D)** Inference retains only the fixed encoder, projection head, and classifier, requiring no evaluation-user calibration data or test-time parameter updates.

GroupNorm (Wu & He, 2018) and ELU. Write the output as $H = F_\theta(x) \in \mathbb{R}^{D \times L}$, where $D = 32$ and L is the output temporal length. This notation does not assume that the convolutions preserve the input length T. We extract the following statistics along the temporal axis:

$$\mu_c(H) = \frac{1}{L}\sum_{t=1}^{L} H_{ct}, \qquad \sigma_c(H) = \max\left\{\sqrt{\frac{1}{L}\sum_{t=1}^{L}\big(H_{ct} - \mu_c(H)\big)^2}, \epsilon\right\}, \tag{1}$$
$$q(H) = [\boldsymbol{\mu}(H); \log \boldsymbol{\sigma}(H)] \in \mathbb{R}^{2D}.$$

Here $c \in \{1, \ldots, D\}$, $\epsilon = 10^{-5}$, and the logarithm is applied elementwise. A projection head g_ϕ with two 128-dimensional fully connected layers produces $e = g_\phi(q(H)) \in \mathbb{R}^{128}$, with ELU and dropout at rate 0.2 between the layers (Srivastava et al., 2014). The normalized representation $z = e/\|e\|_2$ is used for regularization. The identity logits use the unnormalized embedding: $\ell = W_{\text{cls}} e + b_{\text{cls}}$, with $W_{\text{cls}} \in \mathbb{R}^{|\mathcal{Y}| \times 128}$ and $b_{\text{cls}} \in \mathbb{R}^{|\mathcal{Y}|}$. Thus, the classifier is linear rather than cosine-based. GroupNorm does not guarantee removal of all session information.

3.2 Paired Session-Drift Estimation

After source-only warm-up, let θ_0 denote the encoder parameters. For each auxiliary identity $a \in \mathcal{A}$, let $\mathcal{S}_a^{\text{tr}}$ and $\mathcal{T}_a$ contain its source-training and auxiliary target-session segments, respectively. The empirical drift bank is $\mathcal{D} = \{d_a : a \in \mathcal{A}\}$, where

$$d_a = \frac{1}{|\mathcal{T}_a|}\sum_{x \in \mathcal{T}_a} q(F_{\theta_0}(x)) - \frac{1}{|\mathcal{S}_a^{\text{tr}}|}\sum_{x \in \mathcal{S}_a^{\text{tr}}} q(F_{\theta_0}(x)). \tag{2}$$

Pairing is at the identity level: the two session means correspond to the same person, without requiring one-to-one correspondence between individual trials. We construct the bank without gradient tracking and retain each identity's $2D$-dimensional drift separately. Auxiliary target recordings are used only to estimate these statistics; they do not enter the identity cross-entropy term of the full method. The bank remains fixed during subsequent training, while the encoder continues to update. We examine the resulting potential mismatch between the stored drifts and the evolving feature coordinates in the diagnostic experiments.

To clarify what differencing removes, suppose the mean statistical vector for identity a in session s admits an approximate decomposition $\bar{q}_{a,s} = v_a + h_s + \eta_{a,s}$, where v_a is a stable identity component, h_s is a shared session component, and $\eta_{a,s}$ is an interaction residual. Then $\bar{q}_{a,\mathrm{tgt}} - \bar{q}_{a,\mathrm{src}} = (h_{\mathrm{tgt}} - h_{\mathrm{src}}) + (\eta_{a,\mathrm{tgt}} - \eta_{a,\mathrm{src}})$. Differencing cancels v_a but retains the interaction residual. Equation (2) therefore estimates an empirical perturbation, not a strictly disentangled session factor. Its transferability depends on the extent to which session variation is shared across identities.

3.3 Cross-Identity Drift and Representation Consistency

Consider a source-training batch $\{(x_i, y_i)\}_{i=1}^{B}$. For sample i, define the eligible auxiliary identities as $\mathcal{A}_i = \mathcal{A} \setminus \{y_i\}$, which equals $\mathcal{A}$ when $y_i \notin \mathcal{A}$. We sample K drift indices $a_{ik} \in \mathcal{A}_i$ with replacement. Each view independently samples a magnitude $\alpha_{ik} \sim \mathcal{U}(0.5, 1.5)$. Write $d_a = [d_a^{\mu}; d_a^{\log\sigma}]$, where both components belong to $\mathbb{R}^D$; the second component is a shift in log standard deviation, not in standard deviation itself. Let $H_i = F_\theta(x_i)$, $\boldsymbol{\mu}_i = \boldsymbol{\mu}(H_i)$, and $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}(H_i)$. The augmented feature map is

$$\begin{aligned} \widetilde{H}_i^{(k)} &= \exp\bigl(\log \boldsymbol{\sigma}_i + \alpha_{ik} d_{a_{ik}}^{\log\sigma}\bigr) \odot \frac{H_i - \boldsymbol{\mu}_i}{\boldsymbol{\sigma}_i} + \boldsymbol{\mu}_i + \alpha_{ik} d_{a_{ik}}^{\mu} \\ &= \exp\bigl(\alpha_{ik} d_{a_{ik}}^{\log\sigma}\bigr) \odot (H_i - \boldsymbol{\mu}_i) + \boldsymbol{\mu}_i + \alpha_{ik} d_{a_{ik}}^{\mu}. \end{aligned} \tag{3}$$

Exponentiation, multiplication, and division are elementwise, and all statistical vectors are broadcast along time. The second line is algebraically equivalent to the first because every entry of $\boldsymbol{\sigma}_i$ is positive. The transformation shifts the channel means and applies a positive channel-wise scale to the centered temporal waveforms. If the standard-deviation floor is inactive for both H_i and $\widetilde{H}_i^{(k)}$, then $q(\widetilde{H}_i^{(k)}) = q(H_i) + \alpha_{ik} d_{a_{ik}}$; the log-scale identity need not hold when either floor is active. Values of $\alpha_{ik} > 1$ allow extrapolation beyond the estimated drift magnitude.

Set $H_i^{(0)} = H_i$ and $H_i^{(k)} = \widetilde{H}_i^{(k)}$ for $k \geq 1$. For every view, compute $e_i^{(k)} = g_\phi(q(H_i^{(k)}))$ and $z_i^{(k)} = e_i^{(k)} / \|e_i^{(k)}\|_2$. AugMix uses prediction consistency across augmented views to improve image-classification robustness (Hendrycks et al., 2020); here, consistency is imposed on normalized identity representations. With $K = 4$ by default, the cosine-consistency loss is

$$\mathcal{L}_{\mathrm{SI}} = \frac{1}{BK} \sum_{i=1}^{B} \sum_{k=1}^{K} \left(1 - \left\langle z_i^{(0)}, z_i^{(k)} \right\rangle\right). \tag{4}$$

For unit vectors, each summand equals $\frac{1}{2}\|z_i^{(0)} - z_i^{(k)}\|_2^2$. Here, direction refers to the orientation of an embedding, not to the source-to-target transfer direction.

3.4 Worst-View Prototype Regularization

Let $\mathcal{Y}_B = \{y_i : 1 \leq i \leq B\}$ be the identities in the batch, and let $n_c = |\{i : y_i = c\}|$. We construct prototypes from original and augmented representations:

$$u_c = \frac{1}{(K+1)n_c} \sum_{i:y_i=c} \sum_{k=0}^{K} z_i^{(k)}, \qquad p_c = \frac{u_c}{\|u_c\|_2}, \quad c \in \mathcal{Y}_B. \tag{5}$$

The unit-vector definitions assume nonzero embeddings and prototype means. The prototypes participate in backpropagation and include the query sample and all of its augmented views. They therefore act as an in-batch discriminative regularizer, rather than prototypes estimated from an independent support set. With temperature $\tau = 0.07$, we compute the prototype classification loss

for each augmented view and maximize it over the K augmented views of each sample:

$$r_{ik} = -\log \frac{\exp\Big(\langle z_i^{(k)}, p_{y_i}\rangle/\tau\Big)}{\sum_{c\in\mathcal{Y}_B}\exp\Big(\langle z_i^{(k)}, p_c\rangle/\tau\Big)}, \qquad 1 \le k \le K,$$
$$\mathcal{L}_{\mathrm{RP}} = \frac{1}{B}\sum_{i=1}^{B} \max_{1\le k\le K} r_{ik}. \tag{6}$$

The original view contributes to prototype construction but is not included in this maximization. Each view is compared only with the classes present in the batch. If a batch contains only one identity, its prototype classification loss is zero.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{SI}}\mathcal{L}_{\mathrm{SI}} + \lambda_{\mathrm{RP}}\mathcal{L}_{\mathrm{RP}}. \tag{7}$$

Here $\lambda_{\mathrm{SI}} = 1$, $\lambda_{\mathrm{RP}} = 0.5$, and $\mathcal{L}_{\mathrm{CE}} = \frac{1}{B}\sum_{i=1}^{B}\mathrm{CE}_{0.1}(\ell_i^{(0)}, y_i)$, with $\ell_i^{(0)} = W_{\mathrm{cls}}e_i^{(0)} + b_{\mathrm{cls}}$. The notation $\mathrm{CE}_{0.1}$ denotes cross-entropy over all $|\mathcal{Y}|$ identities with label-smoothing coefficient 0.1. Identity cross-entropy is applied only to the original views. The worst-view term follows the robust-optimization intuition of emphasizing difficult perturbations (Madry et al., 2018), but its maximum is over finitely sampled views, not an adversarially optimized uncertainty set. It does not guarantee robustness to every unseen session. At inference, the predicted identity is $\widehat{y}(x) = \arg\max_{c\in\mathcal{Y}} \ell_c(x)$. The encoder, projection head, and classifier are fixed; neither the drift bank nor the prototypes are needed, and no target-batch adaptation is performed.

4 Experimental Setup

Datasets and isolation. MCD contains 23 participants, 60 channels, and 64 trials of 1,600 samples per participant under each of the IQ and memory conditions. Its auxiliary/evaluation split is 15/8; the main direction is IQ→memory, with the reverse reported separately. Task and session changes are confounded. M3CV uses the 95 participants with two sessions in the 64-channel, 250-Hz distribution (Huang et al., 2022), a 20/75 split, and 4-s epochs; its main direction is $S_1 \to S_2$. SEED-IV uses 15 participants, 62 channels, three sessions, a 10/5 split, and non-overlapping 4-s windows of 200-Hz time-domain EEG (Zheng et al., 2019). All six directed transfers are evaluated. Candidate sets remain 23/95/15 classes. Dataset-specific models are trained separately with fixed identity splits across methods and directions.

For MCD, split seed 42 defines identities, and a stratified 70/30 source split gives 1,030/442 training/held-out trials; auxiliary and evaluation targets contain 960 and 512 trials. SEED-IV splits complete movie trials before windowing (16/8 source training/held-out trials per identity) and excludes the unused third session. M3CV preserves original epoch boundaries. Normalization and model selection exclude evaluation-user target data. Task/emotion-stratified sampling and segment lists are shared across methods. Appendix B details provenance and recording-level split requirements.

Training and comparison groups. We use AdamW (Loshchilov & Hutter, 2019), batch size 128, weight decay 0.001, and label smoothing 0.1 (Szegedy et al., 2016): 30 warm-up epochs at learning rate 0.001, then 40 fine-tuning epochs at 0.0005. Same-backbone strategies share warm-up weights within each run and reset the same training seed. External backbones receive 70 epochs; equal epochs do not imply equal compute. Group A in Table 1 uses source-only information; Group B fixes our backbone and auxiliary-target budget. Source+Aux CE adds auxiliary supervision, Deep CORAL (Sun & Saenko, 2016) and DANN (Ganin et al., 2016) align auxiliary domains, RandomShift shuffles drift coordinates, and SHIFT removes RP from the full configuration. Appendix B provides training details.

Five-run reporting. Every experiment reports the arithmetic mean of five runs. Rank-1 is macro-averaged over evaluation identities while retaining all enrolled candidates. We also report Rank-5, one-vs-rest EER and AUC, and the highest TAR at FAR$\le 1\%/0.1\%$ (T1/T0.1). Each metric is computed within each run and direction before averaging. SEED-IV then equally averages its six

Table 1: Identification and verification on three datasets (%). All entries are five-run means. The two panels contain complementary metrics for the same models.

Identification and error rates	MCD			M3CV			SEED-IV		
Model or strategy	R1↑	R5↑	EER↓	R1↑	R5↑	EER↓	R1↑	R5↑	EER↓
A: source-session-only training									
DeepConvNet	54.77	84.06	16.82	62.18	85.81	9.83	72.46	94.40	12.64
ShallowConvNet	57.70	85.43	15.41	66.35	87.56	8.72	73.18	94.54	12.21
EEGNet	56.37	84.84	15.96	64.82	86.96	9.06	73.54	94.72	12.03
EEG-TCNet	60.20	86.68	14.17	68.41	88.48	8.12	76.27	95.54	10.68
FBCNet	62.42	87.89	13.28	70.16	89.34	7.48	77.36	95.98	10.14
TSception	61.21	87.15	13.74	68.93	88.68	7.91	78.62	96.27	9.61
ATCNet	64.34	88.79	12.36	72.48	90.32	6.91	79.48	96.61	9.24
EEG Conformer	63.71	88.40	12.81	71.76	89.92	7.18	80.42	96.84	8.86
EEG-SimpleConv	62.15	87.70	13.42	70.83	89.58	7.41	77.95	96.11	9.98
EEG-Deformer	69.92	91.56	10.18	76.96	92.25	5.51	84.57	98.20	7.46
TCFormer	70.16	91.72	10.07	77.19	92.41	5.42	84.49	**98.24**	7.51
CAGCNet	**70.39**	**91.76**	**9.94**	**77.41**	**92.44**	**5.32**	**84.68**	98.23	**7.39**
Backbone CE	59.18	84.92	14.38	67.42	88.10	8.26	74.63	95.07	11.87
MixStyle	62.11	87.62	13.46	69.73	89.05	7.84	77.21	95.83	10.52
B: same backbone and auxiliary-target budget									
Source+Aux CE	70.23	91.72	10.14	77.38	92.45	5.39	84.63	98.24	7.48
Deep CORAL	69.84	91.48	10.29	77.11	92.28	5.46	84.41	98.12	7.57
DANN	70.31	91.88	9.91	77.52	92.52	5.29	84.79	98.30	7.37
RandomShift	63.91	88.52	13.06	71.84	89.98	7.13	79.38	96.53	9.47
SHIFT	70.66	91.84	9.83	77.61	92.55	5.23	84.87	98.32	7.32
SHIFT+RP	**70.94**	**92.07**	**9.72**	**77.83**	**92.64**	**5.14**	**85.06**	**98.37**	**7.24**

ROC and low-FAR verification	MCD			M3CV			SEED-IV		
Model or strategy	AUC↑	T1↑	T0.1↑	AUC↑	T1↑	T0.1↑	AUC↑	T1↑	T0.1↑
A: source-session-only training									
DeepConvNet	92.55	38.22	13.35	96.14	58.54	30.62	95.33	47.97	22.37
ShallowConvNet	93.37	42.53	16.97	96.82	63.69	36.45	95.55	49.68	23.96
EEGNet	93.06	40.88	15.58	96.62	62.16	34.69	95.67	50.50	24.71
EEG-TCNet	94.10	46.41	20.21	97.20	66.53	39.64	96.41	56.08	29.88
FBCNet	94.64	49.27	22.60	97.62	69.65	43.10	96.73	58.43	32.03
TSception	94.35	47.73	21.32	97.33	67.51	40.74	97.00	60.53	33.99
ATCNet	95.18	52.10	24.97	97.97	72.30	46.10	97.23	62.15	35.47
EEG Conformer	94.90	50.62	23.74	97.78	70.94	44.61	97.42	63.66	36.88
EEG-SimpleConv	94.55	48.79	22.20	97.65	69.93	43.44	96.81	59.05	32.61
EEG-Deformer	96.46	58.86	30.63	98.84	78.86	53.51	98.21	69.55	42.31
TCFormer	96.53	59.27	30.96	98.91	79.35	54.03	98.20	69.40	42.16
CAGCNet	**96.60**	**59.84**	**31.25**	**98.96**	**79.75**	**54.51**	**98.25**	**69.83**	**42.58**
Backbone CE	93.99	44.61	19.71	97.12	65.93	38.94	95.76	51.17	25.33
MixStyle	94.52	48.61	22.06	97.37	67.85	41.12	96.50	56.74	30.49
B: same backbone and auxiliary-target budget									
Source+Aux CE	96.48	59.01	30.75	98.92	79.45	54.16	98.21	69.49	42.25
Deep CORAL	96.39	58.49	30.32	98.87	79.07	53.75	98.15	69.06	41.87
DANN	96.62	59.74	31.36	98.99	79.93	54.70	98.27	69.96	42.68
RandomShift	94.76	49.87	23.11	97.82	71.21	44.90	97.09	61.14	34.55
SHIFT	96.67	60.00	31.56	99.03	80.21	55.01	98.30	70.16	42.87
SHIFT+RP	**96.73**	**60.31**	**31.84**	**99.08**	**80.62**	**55.48**	**98.34**	**70.48**	**43.17**

Note. R1/R5: Rank-1/5; EER: equal error rate; AUC: one-vs-rest ROC area; T1/T0.1: highest TAR at FAR$\leq 1\%/0.1\%$. Boldface marks the best value within each information group. Backbone CE, MixStyle, and Group B share our backbone; Groups A and B have different training-information budgets.

direction means; the supplementary MCD mean averages two directions. Scores are never pooled across runs to construct one ROC curve. Gains are differences between means, not significance claims. Formal definitions and aggregation details appear in Appendix B.4.

5 RESULTS

Architectural and information-matched comparisons. Table 1 reports all six metrics for the three datasets. The full configuration achieves Rank-1 of 70.94/77.83/85.06% on MCD/M3CV/SEED-IV and the lowest EER in the table. Relative to CAGCNet, the strongest external source-only baseline, Rank-1 gains are 0.55/0.42/0.38 percentage points and EER decreases are 0.22/0.18/0.15 points. These comparisons involve different training information and do not isolate our mechanisms. Under the matched backbone and auxiliary budget, the Rank-1 gains over DANN are 0.63/0.31/0.27 points; adding RP to SHIFT yields 0.28/0.22/0.19 points. The 11.76-point MCD gap from source-only Backbone CE also includes the effect of auxiliary data. We emphasize consistent, modest differences rather than universal superiority.

Table 2: Transfer-direction and input-perturbation results (%). Each condition is a five-run mean; directional averages are equally weighted.

		MCD identification/verification					MCD perturbation R1				
Model	Info.	F-R1	R-R1	Avg.	F-R5	TAR	Clean	20 dB	10 dB	C3	C6
EEGNet	S	56.37	59.57	57.97	84.84	40.88	56.37	54.18	48.36	51.56	45.90
TCFormer	S	70.16	72.34	71.25	91.72	59.27	70.16	68.05	61.21	65.04	59.10
CAGCNet	S	70.39	72.19	71.29	91.76	59.84	70.39	68.20	61.33	65.12	59.22
CE	S	59.18	61.80	60.49	84.92	44.61	59.18	57.03	50.47	54.49	48.71
SHIFT	A	70.66	72.42	71.54	91.84	60.00	70.66	68.44	61.52	65.23	59.41
SHIFT+RP	A	**70.94**	**72.70**	**71.82**	**92.07**	**60.31**	**70.94**	**68.59**	**61.68**	**65.43**	**59.57**

		SEED-IV directional R1						
Model	Info.	$1 \to 2$	$1 \to 3$	$2 \to 1$	$2 \to 3$	$3 \to 1$	$3 \to 2$	Avg.
EEGNet	S	74.18	70.59	75.01	74.43	71.68	75.35	73.54
TCFormer	S	84.54	82.30	85.78	85.36	83.05	85.93	84.49
CAGCNet	S	84.71	82.57	85.89	85.55	83.26	86.10	84.68
CE	S	75.26	71.84	76.37	75.61	72.58	76.12	74.63
SHIFT	A	84.94	82.73	86.14	85.67	83.50	86.26	84.87
SHIFT+RP	A	**85.12**	**82.94**	**86.31**	**85.87**	**83.66**	**86.48**	**85.06**

Note. S/A: source-only/auxiliary-target information. F/R: IQ→memory/reverse; TAR: TAR at FAR$\leq 1\%$. C3/C6: zeroing 3/6 channels; 20/10 dB: Gaussian-noise SNR. MCD main results use F; its bidirectional average is supplementary.

Transfer directions and perturbations. Table 2 reports both MCD directions, all six SEED-IV directions, and MCD noise/channel-masking conditions. The full method improves over SHIFT in every evaluated direction; SEED-IV gains range from 0.16 to 0.22 points. Its direction means span 82.94–86.48%, demonstrating residual variation in session difficulty. M3CV reverse transfer is not inferred from the measured direction. Gaussian noise uses each trial's signal power; dropped channels are zeroed after fixed source normalization, with shared seeds and masks. No test-time updates occur. Gains over TCFormer under the four perturbations are 0.39–0.54 points, but different training information prevents attributing these gaps to a single component.

Integrated visualization. Figure 2 relates quantitative performance to spatial sensitivity and representation geometry. Rank-1/EER trends agree with the main comparisons; directional profiles expose variability hidden by aggregate means. Spatial attributions and learned weights are non-uniform, but neither demonstrates physiological identity sources or causal dependence on electrodes (Adebayo et al., 2018). Two-dimensional views suggest differences in identity separation, without recovering high-dimensional classifier boundaries. These visualizations motivate, but cannot replace, controlled ablations. Appendix D provides the full panel-by-panel interpretation.

6 ABLATIONS AND DIAGNOSTICS

Drift structure and objectives. Table 3 retains all mechanism, prototype, drift-update, and hyperparameter results. Removing mean or scale shifts reduces performance, with their relative contributions

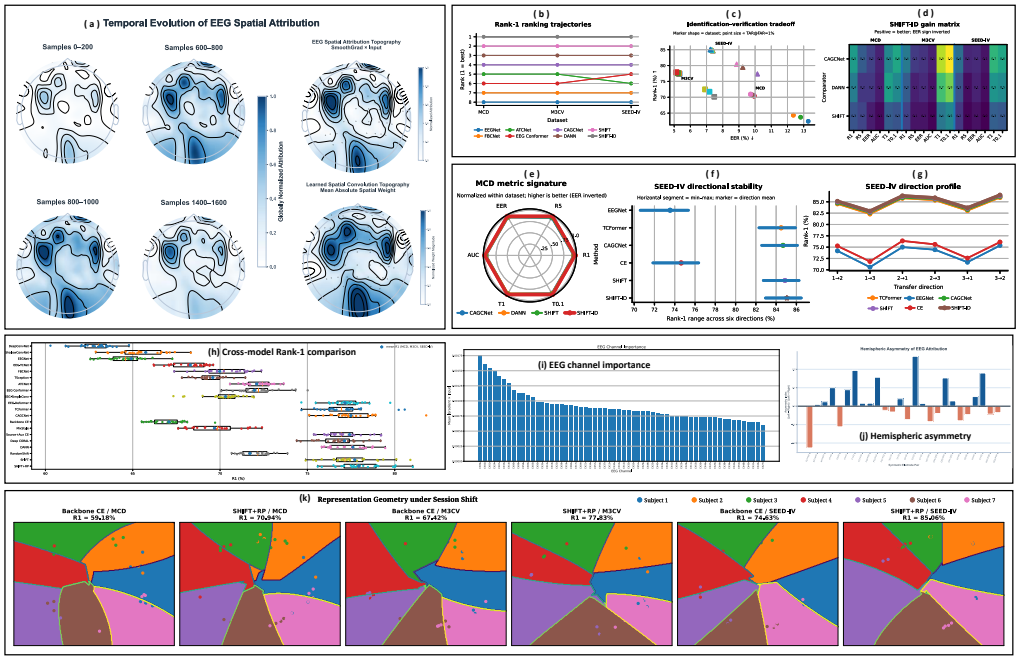


Figure 2: Integrated visualization of cross-session performance, EEG spatial attribution, and representation geometry for SHIFT-ID. (a) EEG spatial attribution at different temporal segments, together with the aggregated SmoothGrad×Input topography and learned spatial-convolution weight map; (b) relative Rank-1 ranking trajectories of representative methods on MCD, M3CV, and SEED-IV; (c) joint identification–verification relationships among Rank-1, EER, and TAR at FAR$\leq 1\%$, where smaller horizontal values and larger vertical values are better and marker size reflects TAR; (d) multi-metric gain matrix of the full configuration relative to CAGCNet, DANN, and SHIFT, with EER sign-adjusted so that a decrease is represented as a positive gain; (e) direction-aligned and normalized multi-metric performance profile on MCD; (f)–(g) range, mean, and per-direction Rank-1 over the six directed transfers of SEED-IV; (h) side-by-side Rank-1 comparison and descriptive summary of different models on the three datasets; (i) EEG channel-importance ranking; (j) left–right attribution differences for symmetric electrode pairs; and (k) representation geometry under session change for source-only CE and the full configuration on MCD, M3CV, and SEED-IV. Quantitative performance summaries use five-run means. The range in panel (f) is across the six direction-specific means, not across runs. Cross-dataset aggregation is used only to facilitate visual comparison of relative model positions and is not defined as a new overall performance metric. Spatial attributions and two-dimensional representations are used to interpret model behavior and are not evidence of physiological causality or statistical significance.

differing by dataset. Unpaired differences and randomized coordinates produce larger losses, supporting the roles of within-identity pairing and coordinate correspondence. The full configuration exceeds average-view aggregation by 0.20/0.16/0.14 points on MCD/M3CV/SEED-IV. These small margins support a limited incremental benefit from selecting difficult views.

Prototype and drift-bank diagnostics. P0–P2 use at least two distinct original trials per identity. P1 excludes the query trial and all its augmented views, while P2 stops prototype gradients. Their lower accuracies show that self-inclusion and gradient flow affect this regularizer, rather than establishing an independent-support interpretation. D0–D2 use default sampling and compare a fixed bank, rebuilding every 10 epochs, and a frozen encoder. The fixed bank performs best among these tested alternatives, while coordinate mismatch during training remains a limitation. Results from the two sampling schemes are not mixed.

Sensitivity, auxiliary budget, and cost. Panel B changes one factor at a time. Its test accuracies describe sensitivity; evaluation target data are excluded from hyperparameter selection. With 5/10/15

Table 3: Component diagnostics and hyperparameter sensitivity. R1 is in percent; Δ is in percentage points.

A Components		Mechanism Ablation										Prototype			Drift Update		
Dataset	Metric	C0	C1	C2	C3	C4	C5	C6	C7	C8	Full	P0	P1	P2	D0	D1	D2
MCD	R1	59.18	70.66	70.08	70.27	70.12	70.43	63.98	65.27	70.74	**70.94**	**70.82**	70.55	70.31	**70.94**	70.70	69.49
M3CV	R1	67.42	77.61	77.34	77.25	77.38	77.42	71.66	73.14	77.67	**77.83**	**77.76**	77.58	77.39	**77.83**	77.68	76.93
SEED-IV	R1	74.63	84.87	84.54	84.49	84.32	84.62	79.72	80.81	84.92	**85.06**	**84.98**	84.83	84.65	**85.06**	84.91	84.02
MCD	Δ	11.76	0.28	0.86	0.67	0.82	0.51	6.96	5.67	0.20	0.00	0.00	0.27	0.51	0.00	0.24	1.45

B Hyperparameter	K				λ_{SI}				λ_{RP}				τ				α Range			
Setting	1	2	4	8	0	0.5	1	2	0	0.25	0.5	1	0.03	0.07	0.1	0.2	$[0, 0.5]$	$[0.5, 1]$	$[0.5, 1.5]$	$[1, 2]$
MCD R1	70.12	70.55	**70.94**	70.70	70.08	70.63	**70.94**	70.39	70.66	70.78	**70.94**	70.47	70.43	**70.94**	70.74	70.20	69.84	70.59	**70.94**	70.16

Note. All R1 values are means over five runs; Δ denotes a difference between these means. In Panel A, Δ is the decrease in R1, in percentage points, relative to the corresponding reference: 70.94 for the mechanism-ablation and drift-update groups, and 70.82 for the prototype group. Results from the two batch-sampling schemes are not mixed. In Panel B, only one factor is changed at a time. Boldface indicates the best result within each group.

MCD auxiliary identities and 64 trials each, R1 is 69.88/70.59/70.94%. At a fixed total of 240 trials, allocations of 5×48, 10×24, and 15×16 yield 69.41/69.92/70.20%. These are within-method budget analyses. The MCD inference model has 61,191 parameters, equal to the same-backbone CE model: drift sampling and prototypes add no inference branch. Equal parameter counts do not imply equal latency. Appendix E gives the complete diagnostic protocol and budget discussion.

7 Discussion and Conclusion

SHIFT-ID transfers paired empirical session differences across identities and preserves discrimination with consistency and worst-view prototype regularization. Across three datasets, gains recur under matched information, multiple transfer directions, and controlled ablations, but absolute improvements over strong baselines are modest. Five-run means alone do not establish statistical significance, and correlated EEG windows are not independent subjects.

Limitations include identity–session interactions retained in empirical drifts, mismatch between a fixed bank and an evolving encoder, and query self-inclusion in prototypes. MCD confounds task and session; M3CV and SEED-IV introduce distinct paradigm and affective-state changes. These experiments establish neither physiological identity invariance nor long-term cross-device stability. The setting remains closed-set recognition with auxiliary histories: it does not cover unseen identities, impostor rejection, replay attacks, or calibrated authentication thresholds. Appendix F expands the scope and data-availability discussion.

AI Use Statement

Generative AI tools were used to assist with English translation and language editing, mathematical typesetting, and manuscript organization. They were not used to formulate research hypotheses or mathematical claims, develop the proposed methodology, design or conduct experiments, generate or process experimental data, implement the proposed method, interpret experimental results, or determine the scientific conclusions of this work. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

References

J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim. Sanity Checks for Saliency Maps. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/294a8ed24b1ad22ec2e7efea049b8737-Abstract.html.

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. *IEEE Transactions on Industrial Informatics*, 19(2):2249–2258, 2023. doi: 10.1109/TII.2022.3197419. URL https://doi.org/10.1109/TII.2022.3197419.

Hamdi Altaheri, Fakhri Karray, and Amir-Hossein Karimi. Temporal convolutional transformer for EEG based motor imagery decoding. *Scientific Reports*, 15:32959, 2025. doi: 10.1038/s41598-025-16219-7. URL https://doi.org/10.1038/s41598-025-16219-7.

P. Campisi and D. La Rocca. Brain waves for automatic biometric-based user recognition. *IEEE Transactions on Information Forensics and Security*, 9(5):782–800, 2014. doi: 10.1109/TIFS.2014.2308640. URL https://doi.org/10.1109/TIFS.2014.2308640.

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. URL https://doi.org/10.1109/TAFFC.2022.3169001.

Yi Ding, Yong Li, Hao Sun, Rui Liu, Chengxuan Tong, Chenyu Liu, Xinliang Zhou, and Cuntai Guan. EEG-Deformer: A dense convolutional transformer for brain-computer interfaces. *IEEE Journal of Biomedical and Health Informatics*, 29(3):1909–1918, 2025. doi: 10.1109/JBHI.2024.3504604. URL https://doi.org/10.1109/JBHI.2024.3504604.

Yassine El Ouahidi, Vincent Gripon, Bastien Pasdeloup, Ghaith Bouallegue, Nicolas Farrugia, and Giulia Lioi. A strong and simple deep learning baseline for BCI MI decoding. arXiv preprint arXiv:2309.07159, 2023. URL https://arxiv.org/abs/2309.07159.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. doi: 10.1016/j.patrec.2005.10.010. URL https://doi.org/10.1016/j.patrec.2005.10.010.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016. URL https://jmlr.org/papers/v17/15-239.html.

D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1912.02781.

G. Huang, Z. Hu, W. Chen, S. Zhang, Z. Liang, L. Li, L. Zhang, and Z. Zhang. M^3CV: A multi-subject, multi-session, and multi-task database for EEG-based biometrics challenge. *NeuroImage*, 264:119666, 2022. doi: 10.1016/j.neuroimage.2022.119666. URL https://doi.org/10.1016/j.neuroimage.2022.119666.

X. Huang and S. Belongie. Arbitrary Style Transfer in Real-Time With Adaptive Instance Normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1501–1510, 2017. URL https://openaccess.thecvf.com/content_iccv_2017/html/Huang_Arbitrary_Style_Transfer_ICCV_2017_paper.html.

Thorir Mar Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2958–2965. IEEE, 2020. doi: 10.1109/SMC42975.2020.9283028. URL https://doi.org/10.1109/SMC42975.2020.9283028.

V. Jayaram, M. Alamgir, Y. Altun, B. Schölkopf, and M. Grosse-Wentrup. Transfer Learning in Brain-Computer Interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31, 2016. doi: 10.1109/MCI.2015.2501545. URL https://doi.org/10.1109/MCI.2015.2501545.

P. Khosla, P. Teterwak, C. Wang, et al. Supervised Contrastive Learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18661–18673, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html.

D. La Rocca et al. Human brain distinctiveness based on EEG spectral coherence connectivity. *IEEE Transactions on Biomedical Engineering*, 61(9):2406–2412, 2014. doi: 10.1109/TBME.2014.2317881. URL https://doi.org/10.1109/TBME.2014.2317881.

V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. 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.

X. Li, Y. Dai, Y. Ge, J. Liu, Y. Shan, and L.-Y. Duan. Uncertainty Modeling for Out-of-Distribution Generalization. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2202.03958.

I. Loshchilov and F. Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

F. Lotte et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, 2018. doi: 10.1088/1741-2552/aab2f2. URL https://doi.org/10.1088/1741-2552/aab2f2.

A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards Deep Learning Models Resistant to Adversarial Attacks. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1706.06083.

E. Maiorana, D. La Rocca, and P. Campisi. On the Permanence of EEG Signals for Biometric Recognition. *IEEE Transactions on Information Forensics and Security*, 11(1):163–175, 2016. doi: 10.1109/TIFS.2015.2481870. URL https://doi.org/10.1109/TIFS.2015.2481870.

Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A. P. Vinod, Seong-Whan Lee, and Cuntai Guan. FBCNet: A multi-view convolutional neural network for brain-computer interface. arXiv preprint arXiv:2104.01233, 2021. URL https://arxiv.org/abs/2104.01233.

Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, 2019. doi: 10.1088/1741-2552/ab260c. URL https://doi.org/10.1088/1741-2552/ab260c.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017. doi: 10.1002/hbm.23730. URL https://doi.org/10.1002/hbm.23730.

J. Snell, K. Swersky, and R. S. Zemel. Prototypical Networks for Few-shot Learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1703.05175.

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. URL https://doi.org/10.1109/TNSRE.2022.3230250.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL https://jmlr.org/papers/v15/srivastava14a.html.

Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *Computer Vision – ECCV 2016 Workshops*, volume 9915 of *Lecture Notes in Computer Science*, pp. 443–450. Springer, 2016. doi: 10.1007/978-3-319-49409-8_35. URL https://doi.org/10.1007/978-3-319-49409-8_35.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html.

Xinran Wang, Xuanyu Jin, Wanzeng Kong, and Fabio Babiloni. CAGCNet: generalized contrastive learning for person identification based on channel aggregated EEG features. *Cognitive Neurodynamics*, 19:141, 2025. doi: 10.1007/s11571-025-10325-y. URL https://doi.org/10.1007/s11571-025-10325-y.

Y. Wu and K. He. Group Normalization. In *European Conference on Computer Vision*, pp. 3–19, 2018. URL https://arxiv.org/abs/1803.08494.

P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu. Transfer Learning: A Riemannian Geometry Framework With Applications to Brain–Computer Interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2018. doi: 10.1109/TBME.2017.2742541. URL https://doi.org/10.1109/TBME.2017.2742541.

H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1710.09412.

W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. 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.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with MixStyle. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2104.02008.

O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş. Adversarial Deep Learning in EEG Biometrics. *IEEE Signal Processing Letters*, 26(5):710–714, 2019. doi: 10.1109/LSP.2019.2906826. URL https://doi.org/10.1109/LSP.2019.2906826.

A Implementation and Mathematical Details

A.1 Statistical Encoding and Network Architecture

For an input $x \in \mathbb{R}^{C\times T}$, channel-wise normalization uses fixed means and standard deviations estimated from the source training set. The encoder F_θ applies a 1×63 temporal convolution, a spatial convolution spanning all C input channels, and a depthwise temporal convolution of length 15 followed by a pointwise convolution. The channel widths are 16, 32, and 32; each block uses GroupNorm (Wu & He, 2018) and ELU. Write the output as $H = F_\theta(x) \in \mathbb{R}^{D\times L}$, where $D = 32$ and L is the output temporal length. This notation does not assume that the convolutions preserve the input length T. We extract the following statistics along the temporal axis:

$$\mu_c(H) = \frac{1}{L}\sum_{t=1}^{L} H_{ct}, \qquad \sigma_c(H) = \max\left\{ \sqrt{\frac{1}{L}\sum_{t=1}^{L}\big(H_{ct} - \mu_c(H)\big)^2}, \epsilon \right\}, \tag{8}$$
$$q(H) = [\boldsymbol{\mu}(H); \log \boldsymbol{\sigma}(H)] \in \mathbb{R}^{2D}.$$

Here $c \in \{1, \dots, D\}$, $\epsilon = 10^{-5}$, and the logarithm is applied elementwise. A projection head g_ϕ with two 128-dimensional fully connected layers produces $e = g_\phi(q(H)) \in \mathbb{R}^{128}$, with ELU and dropout at rate 0.2 between the layers (Srivastava et al., 2014). The normalized representation $z = e/\|e\|_2$ is used for regularization. The identity logits use the unnormalized embedding: $\ell = W_{\text{cls}}e + b_{\text{cls}}$, with $W_{\text{cls}} \in \mathbb{R}^{|\mathcal{Y}|\times 128}$ and $b_{\text{cls}} \in \mathbb{R}^{|\mathcal{Y}|}$. Thus, the classifier is linear rather than cosine-based. GroupNorm does not guarantee removal of all session information.

A.2 Interpretation of Paired Differencing

To clarify what differencing removes, suppose the mean statistical vector for identity a in session s admits an approximate decomposition $\bar{q}_{a,s} = v_a + h_s + \eta_{a,s}$, where v_a is a stable identity component, h_s is a shared session component, and $\eta_{a,s}$ is an interaction residual. Then $\bar{q}_{a,\text{tgt}} - \bar{q}_{a,\text{src}} = (h_{\text{tgt}} - h_{\text{src}}) + (\eta_{a,\text{tgt}} - \eta_{a,\text{src}})$. Differencing cancels v_a but retains the interaction residual. Equation (2) therefore estimates an empirical perturbation, not a strictly disentangled session factor. Its transferability depends on the extent to which session variation is shared across identities.

A.3 Equivalent Feature Transformation and Boundary Cases

The affine transformation in Equation (3) can also be written as

$$\widetilde{H}_i^{(k)} = \exp\big(\log \boldsymbol{\sigma}_i + \alpha_{ik} d_{a_{ik}}^{\log\sigma}\big) \odot \frac{H_i - \boldsymbol{\mu}_i}{\boldsymbol{\sigma}_i} + \boldsymbol{\mu}_i + \alpha_{ik} d_{a_{ik}}^{\mu}. \tag{9}$$

Because every entry of the floored standard deviation $\boldsymbol{\sigma}_i$ is positive, cancellation gives the main-text expression exactly. If neither the original nor transformed standard deviation activates the floor, $q(\widetilde{H}_i^{(k)}) = q(H_i) + \alpha_{ik} d_{a_{ik}}$. This log-scale identity can fail when either floor is active. Values of $\alpha_{ik} > 1$ extrapolate beyond an observed drift magnitude.

For unit vectors, each term in Equation (4) equals $\frac{1}{2}\|z_i^{(0)} - z_i^{(k)}\|_2^2$. Direction here means the orientation of an embedding, not a source-to-target transfer direction. Unit-vector definitions require nonzero embeddings and prototype means. Each prototype loss compares only identities present in the batch; a batch containing one identity has zero prototype classification loss. Original views contribute to the prototypes but are excluded from the worst-view maximum. The finite sampled maximum does not provide a guarantee over every unseen session or an adversarially optimized uncertainty set.

B Data, Training, and Evaluation Details

B.1 Datasets and Transfer Settings

MCD. MCD is an EEG dataset collected by our research team and contains recordings from 23 participants under two cognitive-task conditions, IQ and memory. All participants were informed

about the study and consented before data collection. Under each condition, the data have dimensions $23 \times 60 \times 1600 \times 64$, corresponding to participants, channels, time samples, and trials per participant, for a total of 1,472 trials per condition. We treat the IQ and memory recordings as two cross-condition sessions and evaluate both IQ→memory and memory→IQ. Because the two conditions also differ in cognitive task, the results reflect both session and task-state differences and are not interpreted as pure temporal session drift.

M3CV. The original study reports 106 participants, of whom 95 completed sessions on two different dates across six paradigms (Huang et al., 2022). The corresponding NEMAR distribution provides 64-channel, 250-Hz data for these 95 participants, with preprocessed epochs of 4 s (1,000 time samples). We construct a 95-class closed-set protocol that excludes unregistered impostors and do not assume that all 106 participants have complete two-session data. A predefined research split of 20 auxiliary and 75 evaluation identities is used; this split is neither an official leaderboard protocol nor a claim about the composition of an official test set.

SEED-IV. SEED-IV contains 15 participants recorded on three different dates. Each session includes 24 movie trials spanning four emotion categories and uses 62-channel EEG (Zheng et al., 2019). Participant IDs are treated as identity labels; emotion labels are used only for balancing and diagnostics, not as a four-class prediction target. Following the official description, we use the time-domain EEG signals downsampled to 200 Hz with non-overlapping 4-s windows of 800 samples. The precomputed differential-entropy features have shape $62 \times N_{\text{win}} \times 5$, where N_{win} is the number of windows in a trial; these features are not used as raw EEG input. All six directed transfers among the three sessions are trained and evaluated separately.

The candidate spaces contain 23, 95, and 15 identities for MCD, M3CV, and SEED-IV, respectively, with auxiliary/evaluation splits of 15/8, 20/75, and 10/5. Within a dataset, every direction and every method uses the same identity split. Datasets are trained independently; identity labels are not merged across datasets, and these experiments are not described as cross-dataset transfer.

The multi-paradigm recordings in M3CV and emotion recordings in SEED-IV introduce different state compositions. The main protocol retains the valid task sets of each dataset and performs within-identity stratified sampling by task or emotion so that states with more samples do not dominate training. All methods use exactly the same segment lists, task encoding, missing-label handling, and quality-exclusion rules, ensuring consistent data conditions across methods.

B.2 Identity Isolation and Recording-Level Splits

For MCD, random seed 42 defines 15 auxiliary identities and 8 evaluation identities. The source session from all 23 participants is split 70/30 in a stratified manner by identity, yielding 1,030 training trials and 442 held-out trials. The 960 target trials from the 15 auxiliary identities are used to build the drift bank, whereas the 512 target trials from the 8 evaluation identities are used only for testing. When multiple trials can be traced to the same continuous recording, recording-level grouping is required to prevent leakage between training and validation.

For SEED-IV, grouping is performed by movie trial before windowing. For each source session and emotion category, four complete trials per participant are used for training and two are held out, yielding 16/8 trials per participant. Windows from the same movie trial are not randomly split across source training and validation. The third session not involved in a given transfer direction is completely excluded, and target auxiliary identities are disjoint from target evaluation identities. No window from an evaluation identity in the target session may be used for normalization, early stopping, channel selection, or tuning the augmentation magnitude.

For M3CV, identity correspondence and segment provenance across the two sessions are verified from the distribution metadata, and only the closed-set portion with valid evaluation labels is scored. Reconstructed continuous files may be concatenations of existing epochs and therefore must not be re-windowed across concatenation boundaries (Huang et al., 2022). If original run information cannot be recovered, we do not report random-window within-session validation; instead, we use a predefined configuration and the last-epoch model. Identities in portions with unpublished labels are not filled in by inference or guesswork.

B.3 Unified Training Budget and Fair Comparison

SHIFT-ID uses AdamW (Loshchilov & Hutter, 2019) with batch size 128, weight decay 0.001, and label smoothing of 0.1 (Szegedy et al., 2016). The source warm-up lasts 30 epochs at learning rate 0.001, followed by 40 fine-tuning epochs at learning rate 0.0005. Every experimental configuration is evaluated over five runs. Within each run, strategies using the same backbone share the warm-up weights, and the same training seed is reset before training each strategy. The MCD identity-partition seed of 42 specifies the data split, separately from the repeated training runs. Each dataset retains its native number of channels and input samples; only the spatial-convolution span and identity-output dimension are adjusted. Temporal kernel sizes are fixed in samples, so we do not claim identical physical time scales across datasets.

External backbones use the same total budget of 70 epochs. This matches the epoch budget but does not imply equal training FLOPs or wall-clock time. Source+Aux CE uses paired auxiliary batches to avoid increasing the number of update steps. Deep CORAL (Sun & Saenko, 2016) and DANN (Ganin et al., 2016) align only source/target recordings of auxiliary identities, while source classification supervision still covers all identities. FBCNet is configured with frequency bands and temporal scales appropriate to each dataset's sampling setting, and TSception constructs its left/right branches according to the channel spatial layout. All comparisons use the same data splits, total epoch budget, and evaluation protocol; access to auxiliary target-session data differs between comparison groups.

B.4 Metrics and Numerical Reporting

Within each run and transfer direction, the primary metric is Rank-1 macro-averaged over evaluation identities:

$$\mathrm{R1} = \frac{100}{|\mathcal{E}|} \sum_{y \in \mathcal{E}} \frac{1}{N_y} \sum_{j=1}^{N_y} \mathbf{1}\{\widehat{y}(x_{yj}) = y\}\,. \tag{10}$$

Here x_{yj} is the jth test sample of identity y, N_y is its test-sample count, and $\widehat{y}$ is the prediction over the full candidate set $\mathcal{Y}$. Rank-5 also retains all candidate identities. For MCD, with 64 trials per evaluation identity, macro-averaged and overall trial-level accuracies coincide within each run. The reported MCD accuracies are means over five runs, rather than the accuracy of a single set of predictions. For SEED-IV, identity-wise results are first averaged within each transfer direction and then equally averaged over the six directions; windows from movie trials of different durations are not pooled across identities and directions for scoring. Verification-style metrics provide supplementary diagnostics using one-vs-rest ROC curves from softmax scores over all candidates (Fawcett, 2006): EER is the error rate at the threshold where the false acceptance rate (FAR) equals the false rejection rate (FRR); AUC is the area under the ROC curve; and T1/T0.1 are the highest true acceptance rates (TAR) at FAR no greater than 1%/0.1%, respectively. For SEED-IV, these metrics are computed independently for each direction and then equally averaged, rather than pooling scores from different direction-specific models. They do not represent calibrated deployment thresholds and cannot be inferred from Rank-1 or EER alone. The reported verification metrics are computed from the corresponding evaluation scores and expressed as percentages to two decimal places. For MCD, IQ→memory is the primary reported direction, with memory→IQ and the bidirectional mean used as supplementary stability analyses.

All experiments, including architectural and training-strategy comparisons, transfer-direction analyses, ablations, hyperparameter sweeps, auxiliary-data budget analyses, and input-perturbation evaluations, report arithmetic means over five runs. Each metric is first computed separately for each run and transfer direction. For a metric m, let $m_d^{(r)}$ denote its value in run r and direction d. We report

$$\overline{m}_d = \frac{1}{5} \sum_{r=1}^{5} m_d^{(r)}, \qquad \overline{m}_{\mathcal{V}} = \frac{1}{|\mathcal{V}|} \sum_{d \in \mathcal{V}} \overline{m}_d, \tag{11}$$

where $\mathcal{V}$ is the set of evaluated transfer directions when a directional summary is reported. For SEED-IV, $|\mathcal{V}| = 6$, so the overall result is an equally weighted mean over six directions and five runs per direction. The supplementary MCD bidirectional summary uses $|\mathcal{V}| = 2$; the main MCD result uses only IQ→memory, and M3CV uses only $S_1 \to S_2$. EER, AUC, T1, and T0.1 are computed from each run's evaluation scores before averaging; scores from different runs are not pooled to construct

a single ROC curve. Reported gains are differences between the corresponding five-run means. Performance values are rounded to two decimal places after averaging. Because some absolute differences between methods are small, we interpret the results as performance comparisons under the current protocol rather than equating sub-percentage-point gains directly with statistical significance. For windowed EEG, statistical interpretation remains at the identity and transfer-direction level; correlated windows are not treated as independent subjects.

C Extended Related Work

C.1 General EEG Models and Identity Representation Learning

EEG decoding has evolved from conventional classifiers to a broad range of deep representation-learning architectures (Lotte et al., 2018; Roy et al., 2019). DeepConvNet and ShallowConvNet extract spatiotemporal information through hierarchical convolution and shallow power modeling, respectively (Schirrmeister et al., 2017); EEGNet provides a parameter-efficient baseline with compact spatiotemporal convolutions (Lawhern et al., 2018). EEG-TCNet emphasizes temporal dependencies (Ingolfsson et al., 2020), FBCNet models spectral structure using filter banks and variance aggregation (Mane et al., 2021), and TSception combines multi-scale temporal convolutions with spatial asymmetry (Ding et al., 2023). Because these models were originally developed mostly for motor-imagery or emotion decoding, their performance on identity recognition must be reevaluated rather than inferred from accuracies reported for their original tasks.

Recent architectures have further expanded the design space by combining convolution and attention. ATCNet integrates attention with temporal convolution (Altaheri et al., 2023); EEG Conformer encodes local patterns convolutionally and then captures long-range dependencies with self-attention (Song et al., 2023); EEG-Deformer introduces information fusion across coarse and fine temporal hierarchies (Ding et al., 2025); and TCFormer combines multi-kernel convolution, grouped-query attention, and temporal-convolution heads (Altaheri et al., 2025). EEG-SimpleConv also demonstrates that standard convolution remains a strong baseline (El Ouahidi et al., 2023). Architectural complexity does not necessarily imply better cross-session stability, particularly when training budgets and input adaptations differ.

For identity recognition specifically, Özdenizci et al. studied adversarial learning for cross-session EEG biometrics (Özdenizci et al., 2019), while CAGCNet learns identity representations through channel aggregation, feature enhancement, and generalized contrastive learning (Wang et al., 2025). Such methods are more directly aligned with our task and indicate that evaluation should not be restricted to older general-purpose EEG networks. We therefore compare both identity-oriented methods and general EEG backbones, and separately examine training strategies under the same auxiliary-data budget.

C.2 Statistical Augmentation and Positioning of This Work

AdaIN performs style transfer by adjusting feature means and variances (Huang & Belongie, 2017), providing a foundation for manipulating feature statistics. MixStyle increases training-distribution diversity by mixing feature statistics (Zhou et al., 2021), while DSU simulates potential distribution shifts through uncertainty modeling of statistics (Li et al., 2022). These operations do not explicitly exploit paired cross-session differences within the same identity and therefore need not preserve realistic directions of session variation. SHIFT-ID first computes cross-session differences for each auxiliary identity and then transfers those differences to other identities while preserving the internal correspondence between statistical coordinates. The goal is to constrain the augmentation range using observable variation, not to claim recovery of a pure or causally interpretable session factor. Unlike mixup, which interpolates both samples and labels (Zhang et al., 2018), our method retains the identity label of the current sample and transfers only the statistical difference.

Prototypical networks organize discriminative space around class centers (Snell et al., 2017), providing a complementary mechanism for stability regularization. Supervised contrastive learning also uses class labels to promote intra-class compactness and inter-class separation (Khosla et al., 2020), but its sample-to-sample contrastive objective differs from the prototype classification regularizer used here. Our prototypes are constructed only within training batches and constrain the hardest classification cases among a finite set of perturbed views; they neither serve as enrollment templates at test time

nor constitute few-shot recognition with an independent support set. The method combines paired drift estimation, cross-identity transfer, and hard-view discrimination. We examine these components through pairing disruption, drift randomization, and alternative loss aggregation.

D Extended Interpretation of the Integrated Visualization

The following discussion refers to Figure 2 in the main text.

Consistency across metrics and datasets. Figure 2(b)–(e) reorganizes the results in Table 1 from several evaluation perspectives. In the Rank-1 ranking trajectories, the full configuration SHIFT+RPranks first on MCD, M3CV, and SEED-IV, with Rank-1 accuracies of 70.94%, 77.83%, and 85.06%, respectively. This pattern shows that the observed difference is not confined to a single dataset. The joint identification–verification plot further shows that, within each dataset, SHIFT+RPlies in the favorable region of higher Rank-1 and lower EER while maintaining high TAR at FAR$\leq 1\%$. Thus, closed-set identification and one-vs-rest verification metrics do not exhibit an obvious conflict. This is consistent with the main table, where SHIFT+RPachieves the best Rank-1 and EER on all three datasets.

Figure 2(d) expands the relative gains to six metrics: Rank-1, Rank-5, EER, AUC, T1, and T0.1. Using the full configuration as reference, the principal cells remain positive relative to CAGCNet, DANN, and SHIFT, which removes the worst-view prototype term. Compared with the strongest external baseline CAGCNet, Rank-1 improves by 0.55, 0.42, and 0.38 percentage points on the three datasets. Compared with DANN under the same backbone and auxiliary budget, the gains are 0.63, 0.31, and 0.27 points; compared with SHIFT, they are 0.28, 0.22, and 0.19 points. These differences remain small in absolute magnitude, so the gain matrix is primarily used to examine whether improvements recur across metrics while color intensity encodes the plotted numerical differences and does not establish statistical significance. The normalized MCD profile likewise shows that the change in the full method is not confined to Rank-1 but also appears in retrieval, ROC, and low-FAR verification metrics. This matches our evaluation principle of emphasizing consistent improvements across conditions rather than a large increase in any single number.

Figure 2(h) places Rank-1 results from the three datasets on a common horizontal axis to inspect the relative positions of different architectures as data conditions change. Some source-only models show substantial performance variation across datasets, whereas SHIFT and SHIFT+RP, which use auxiliary target-session information, generally occupy a higher Rank-1 range. Importantly, the datasets differ in number of identities, session composition, and task state. The cross-dataset average in the figure is therefore only a compact descriptive summary; formal comparisons in this paper remain based on metrics computed separately for each dataset, and the mean Rank-1 across all three datasets is not defined as a new overall accuracy measure.

Stability across transfer directions. Figure 2(f)–(g) further expands the six directed transfers of SEED-IV. The Rank-1 values of the full configuration on $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 1$, $2 \rightarrow 3$, $3 \rightarrow 1$, and $3 \rightarrow 2$ are 85.12%, 82.94%, 86.31%, 85.87%, 83.66%, and 86.48%, respectively, where each directional value is a five-run mean, with an equally weighted six-direction mean of 85.06%. The lowest and highest direction-specific means are 82.94% and 86.48%, showing that different session pairs retain non-negligible differences in difficulty. We therefore do not interpret the between-direction variation as evidence that a completely session-invariant representation has been obtained.

Relative to SHIFT, SHIFT+RPimproves the six directions by 0.18, 0.21, 0.17, 0.20, 0.16, and 0.22 percentage points, respectively. All evaluated directions therefore show improvements with the same sign rather than relying on a few favorable directions to raise the average. The minimum–maximum range and direction mean in Figure 2(f), together with the per-direction curves in Figure 2(g), display this behavior at aggregate and local scales. The results support a stable but limited additional benefit from the worst-view prototype term over these six transfer conditions, while remaining insufficient to extrapolate to arbitrary unseen future sessions.

EEG spatial attribution and channel selectivity. Beyond performance metrics, panels (a), (i), and (j) of Figure 2 examine which EEG information the model uses in input space. Figure 2(a)

shows spatial attribution maps for different temporal segments of the same recording. Variation in high-attribution regions across segments suggests that model sensitivity depends on local EEG content. At the same time, several spatial locations recur across segments and in the aggregated attribution map, suggesting a degree of stable spatial preference. The learned spatial-convolution weights also exhibit a non-uniform spatial pattern, indicating non-uniform learned spatial weighting; weight magnitude alone does not quantify each electrode's contribution to a prediction.

This selectivity is more directly visible in the channel-importance ranking of Figure 2(i). The reported attribution scores decrease gradually across channels, suggesting distributed model sensitivity. Establishing whether accurate recognition actually requires multiple electrodes would require controlled channel-ablation experiments; an attribution ranking alone cannot establish functional dependence. Figure 2(j) further compares attribution differences across symmetric electrode pairs. Both positive and negative differences occur, with clear variation in magnitude, indicating location-dependent left–right asymmetry without supporting the simple conclusion that one hemisphere globally dominates.

SmoothGrad$\times$Input, spatial-convolution weights, and channel importance reflect the sensitivity and parameter use of the trained model; they do not directly establish that specific brain regions are physiological sources of identity. Sanity-check studies of saliency maps have likewise shown that visually structured attribution maps alone are insufficient to establish faithfulness to the learned model (Adebayo et al., 2018). This caveat is particularly important for MCD, where session change is accompanied by a cognitive-task change, while M3CV and SEED-IV also contain paradigm or emotion-state differences. The scalp topographies are therefore interpreted only as model-behavior explanations, not as neurophysiological causal conclusions.

Representation geometry under session changes. Finally, Figure 2(k) compares the two-dimensional representation structures of source-only CE and SHIFT+RPon all three datasets. Rank-1 rises from 59.18% to 70.94% on MCD, from 67.42% to 77.83% on M3CV, and from 74.63% to 85.06% on SEED-IV. Relative to source-only CE, target-session samples under the full configuration exhibit a more consistent correspondence with identity-discriminative regions, with apparently less overlap between some identity clusters. Distances and apparent boundaries in these projections should not be interpreted as the decision boundaries of the original high-dimensional classifier.

This observation is directionally consistent with the training objective. Equation (3) first uses empirical drifts from auxiliary identities to construct training views grounded in observed session variation. $\mathcal{L}_{\mathrm{SI}}$ encourages cosine consistency between original and drifted representations, while $\mathcal{L}_{\mathrm{RP}}$ constrains the most difficult prototype-classification case among sampled views. The method therefore does not require features to remain numerically identical across sessions; instead, it allows their statistical locations and scales to change while attempting to preserve identity-relevant representational relationships. Figure 2(k) provides an intuitive two-dimensional projection of this objective. However, any two-dimensional mapping necessarily compresses high-dimensional distances and local-neighborhood relationships, so it serves only as supporting interpretation and cannot replace Rank-1, EER, or the controlled ablations as quantitative evidence for the proposed mechanisms.

Taken together, Figure 2 shows three mutually consistent observations for the full configuration. Improvements appear across multiple datasets, metrics, and transfer directions; attribution scores suggest non-uniform spatial sensitivity; and the two-dimensional projections suggest clearer identity separation under session changes. These observations are aligned with the intended roles of empirical drift, cosine consistency, and worst-view prototype regularization. The integrated visualization, however, cannot isolate the independent contribution of each component. We therefore next perform controlled experiments that remove drift components, break pairing, modify consistency and prototype constraints, and alter the drift-bank update strategy.

E Additional Diagnostic and Budget Analyses

E.1 Prototype Self-Inclusion and Fixed Drift Bank

The prototype diagnostics in Table 3, Panel A, examine the potential effect of query–prototype self-similarity. The prototype-diagnostic group uniformly uses batch sampling with at least two distinct original trials for each identity. The leave-one-trial-out prototype excludes the query trial together

with all of its augmented views, whereas the stop-gradient configuration blocks backpropagation through the prototype branch. The drift-diagnostic group keeps the default sampling strategy and compares a fixed bank, reconstruction every 10 epochs, and a frozen convolutional encoder. This separation avoids confounding a change in batch sampling with a change in the drift-update strategy.

E.2 Hyperparameters and Auxiliary-Data Budget

Table 3, Panel B, reports the parameter settings and their corresponding Rank-1 accuracies. The zero-consistency-weight and zero-prototype-weight settings correspond to the matching ablations in Panel A. The default configuration has the highest reported MCD test Rank-1 among the settings shown. These test results describe sensitivity and are not a hyperparameter-selection criterion; hyperparameter selection excludes target-session data from evaluation identities.

For the MCD auxiliary-data budget analysis, the eight evaluation identities and all source-session supervision from the 23 available identities are kept unchanged. With target recordings from 5, 10, or all 15 members of the fixed auxiliary pool and 64 target trials per selected identity, the total target-session budgets are 320, 640, and 960 trials, respectively. The corresponding Rank-1 accuracies are 69.88%, 70.59%, and 70.94%, which are 1.06, 0.35, and 0.00 percentage points below the full-budget result. We further fix the total number of auxiliary target-session trials to 240 and use 5×48, 10×24, and 15×16 identity–trial allocations. These settings obtain 69.41%, 69.92%, and 70.20% Rank-1 accuracy, respectively, corresponding to decreases of 1.53, 1.02, and 0.74 points relative to the full budget. These results characterize the behavior of the same method under different auxiliary-data budgets and do not constitute equal-budget comparisons with external algorithms.

E.3 Input Perturbations and Inference Cost

The five rightmost columns of Table 2 report the unperturbed condition, two additive Gaussian-noise conditions, and two channel-masking conditions for fixed models. The 20-dB and 10-dB noise conditions specify signal-to-noise ratios relative to the power of each individual trial, and dropped channels are set to zero after applying the fixed source-domain normalization. All methods share the same perturbation seeds and channel masks, and no parameters are updated at test time. Under the four additional perturbation conditions, the complete method outperforms TCFormer by 0.39–0.54 percentage points. Because the two methods use different training information, these differences are used only to describe overall robustness and are not attributed to any single regularization term.

The training-time drift samples and prototype branch do not enter the inference graph, so the complete method has the same number of inference-time parameters as the CE baseline with the same backbone. Under the MCD configuration, the current architecture contains 61,191 parameters. For the other datasets, the parameter count changes with the channel span and the dimensionality of the identity head. Equal parameter counts do not necessarily imply identical end-to-end latency; accordingly, our complexity claim is limited to the fact that drift sampling and prototype computation introduce no additional inference branch.

F Scope, Limitations, and Data Availability

The proposed framework also has several limitations. First, within-identity differencing does not guarantee complete removal of identity–state interaction effects. Second, a drift bank constructed after a fixed warm-up stage may become progressively mismatched as the feature space continues to evolve. Third, prototype self-inclusion can introduce a self-similarity advantage if it is not carefully controlled. In addition, task and session effects cannot be fully disentangled in MCD, and the multi-paradigm variation in M3CV and affective-state variation in SEED-IV represent distinct sources of distribution shift. Evaluating more datasets broadens the range of conditions considered, but does not by itself establish long-term stability, cross-device generalization, or physiological identity invariance.

Finally, this work considers only closed-set identification among enrolled identities. Historical target-session recordings from auxiliary identities are available during training, whereas target-session recordings from evaluation identities are excluded from all training, normalization, model-selection, and test-time adaptation procedures. The resulting setting is therefore neither unseen-

identity recognition nor open-set authentication. It also does not address impostor rejection, replay attacks, or authentication-threshold calibration. Accordingly, the present results support closed-set cross-session identification under an auxiliary-history setting and should not be extrapolated to open-set authentication, long-term cross-device stability, or universal robustness to arbitrary unseen session conditions.

DATA AVAILABILITY

The MCD dataset used in this study was collected and organized by our research team and is publicly available in a Hugging Face repository under an anonymous account at `https://huggingface.co/datasets/anonymous-588/MCD-EEG`. M3CV and SEED-IV were obtained under the access procedures and usage licenses specified by their original providers. Section 4 and Appendix B describe the data splits, preprocessing procedures, and evaluation protocol.