ORGANIZE ONCE, ADAPT EFFICIENTLY: TARGET-CONDITIONED HYPERGRAPH RETRIEVAL FOR FEW-SHOT MOTOR IMAGERY EEG ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Rapid deployment of motor imagery–based brain–computer interfaces (BCIs) requires reliable decoding for new users with few calibration trials. Prior source-selection methods reuse historical electroencephalography (EEG) data through subject-level screening, trial-level scoring, or pairwise graph propagation. However, exploiting higher-order relations among trials for budget-constrained, target-conditioned source-trial retrieval remains challenging. To address this, we view transfer utility not as an intrinsic pointwise property but as a target-dependent structural quantity jointly determined by local discriminative coherence and cross-domain reproducibility, and propose a neural retrieval framework over a reusable dual-relation source hypergraph. Following an organize-once, retrieve-per-target paradigm, we build this hypergraph offline with local semantic and cross-domain consensus relations. Online, we query this hypergraph with a small target support set to construct a subhypergraph and select transferable trials through structural propagation and budget-constrained ranking. Cross-subject and cross-dataset experiments on BNCI2014-001, Cho2017, and Lee2019-MI show that, in some settings, our method matches or outperforms the All-source training reference while using as little as 2.35% of the available source trials for target adaptation.

1 INTRODUCTION

Rapid deployment of motor imagery (MI) brain–computer interfaces (BCIs) requires reliable electroencephalography (EEG) decoding for new users from only a few calibration trials (Gwon & Ahn, 2024; Wang et al., 2025b). As illustrated in Figure 1, lengthy subject-specific calibration remains a practical bottleneck for scalable BCI research and commercial deployment, motivating adaptation from only limited target-user data. Large amounts of labeled EEG data from historical users and sessions provide a valuable resource for reducing this calibration burden. However, EEG signals vary substantially across users, sessions, and acquisition conditions (Ng & Guan, 2024), making source trials unequally transferable to a new target user (Huang et al., 2026). The key challenge is therefore not simply how to use more source data, but how to identify the source trials that are most useful for each target under a limited data budget.

Existing approaches reuse source EEG through full-source pooling (El Ouahidi et al., 2024), subject-level screening (Zhu et al., 2024), or trial-level selection (Zheng & Lin, 2024; Luo, 2024) (Figure 2). Full-source pooling performs no source selection, subject-level screening operates at a coarse granularity, and trial-level methods typically rely on pointwise scores or pairwise relations. Despite these differences, existing approaches remain limited in modeling higher-order source structures that may jointly determine transferability to a specific target.

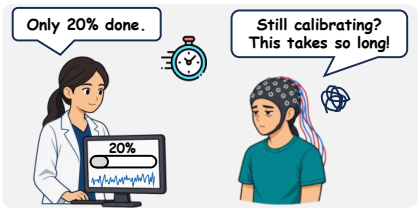


Figure 1: BCI calibration burden.

We argue that source-trial transferability is not an intrinsic sample property, but a target-dependent structural utility. A useful source trial should participate in a locally coherent discriminative pattern

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.

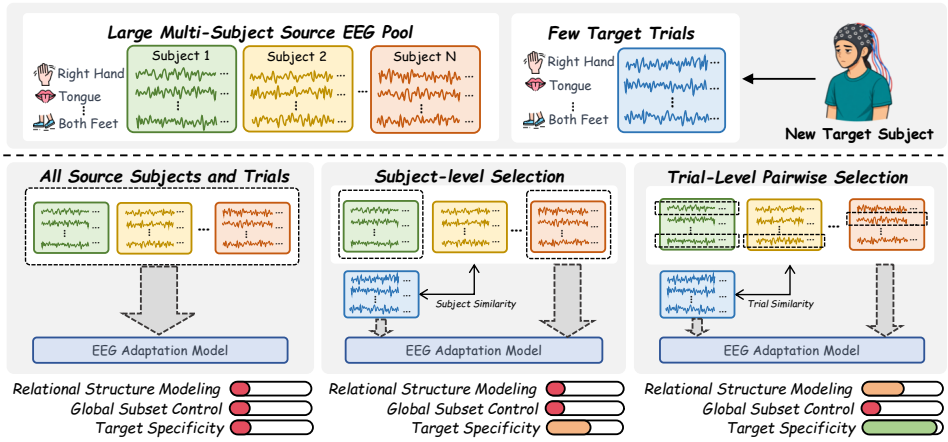


Figure 2: Comparison of source-data utilization for few-shot motor-imagery EEG adaptation.

that also recurs across source domains, providing complementary evidence of local discriminative coherence and cross-domain reproducibility. This motivates an organize-once, retrieve-per-target paradigm: source trials are first organized into reusable higher-order structures, from which we retrieve only the patterns relevant to the current task using a small target support set.

Based on this view, we propose Target-Conditioned EEG Transfer via Hypergraph-Guided Efficient Retrieval (TETHER). Offline, TETHER organizes source trials through complementary hypergraph relations capturing local coherence and cross-domain consensus, and learns the retriever through source-only pseudo-target episodes supervised by downstream adaptation performance. Online, TETHER uses a few labeled target trials to query this reusable structure and retrieve a fixed-budget source subset for adaptation. Extensive within-dataset cross-subject and cross-dataset experiments on three public MI datasets show that TETHER consistently outperforms budget-matched baselines across target-shot settings while providing strong data and computational efficiency. Our contributions are threefold:

- We reformulate few-shot EEG adaptation as target-conditioned source-trial retrieval, modeling transferability as target-dependent structural utility.
- We propose TETHER, a target-conditioned retriever built on a reusable dual-relation source hypergraph and trained through utility-supervised retriever learning.
- Experiments on three MI datasets show consistent gains with only 2.35% of source trials and sub-second online adaptation.

2 Related Work

Cross-Subject Transfer Learning for Brain–Computer Interfaces. Cross-subject transfer learning reduces calibration for new users by leveraging data or knowledge from historical subjects and sessions (Zhang et al., 2024). Existing approaches broadly follow data alignment, transferable representation learning, and target or model adaptation. Alignment methods reduce inter-subject distribution shifts through transformations such as Euclidean alignment (Wu, 2025; Wang et al., 2025c), while representation-learning approaches seek subject-invariant features through contrastive learning, self-attention, and large-scale pretraining across heterogeneous EEG datasets (Sun et al., 2024; Zhi et al., 2025; Yang et al., 2023; Jiang et al., 2024; Wang et al., 2024; 2025a). Adaptation-based strategies further transfer source knowledge toward unseen users through generative, subject-adaptive, or pseudo-label-based mechanisms (An et al., 2024; Yang et al., 2025). Despite improved cross-subject decoding, these methods primarily focus on how source knowledge is aligned or adapted, rather than which source trials should be retained for a specific target under a limited data budget.

Source-Domain and Source-Trial Selection. Source selection has been studied to mitigate negative transfer caused by mismatched subjects or weakly transferable trials. At the domain level, recent methods select source subjects according to distribution similarity or class-discriminative Riemannian criteria (Jin et al., 2024; Ding et al., 2026). At the trial level, Manifold Trial Selection (MTS) and Manifold Embedded Instance Selection (MEIS) estimate trial transferability in manifold space (Liang et al., 2021; 2024), while GITGAN and curriculum-transfer methods further suppress source outliers or jointly consider domain- and sample-level transferability (Yin et al., 2024; Gao et al., 2026). Beyond EEG, recent work has explored compute-constrained and target-aware data selection (Yin & Rush, 2025; Jain et al., 2026; Das et al., 2026). Within EEG, however, transferability is still predominantly assessed through domain-level statistics or individual-trial scores, with limited modeling of interactions among multiple source trials.

Graph and Hypergraph Modeling for EEG. Graph neural networks model non-Euclidean EEG structure through pairwise relations among electrodes or brain regions (Ding et al., 2024; Shi et al., 2025; Zhang et al., 2025a). Hypergraphs extend pairwise graphs to higher-order relations among multiple nodes (Antelmi et al., 2023; Chien et al., 2022; Yan et al., 2024; Tang et al., 2025). Recent EEG studies have applied hypergraph learning to multimodal EEG–fMRI modeling, speech-imagery decoding, and emotion recognition (Zhang et al., 2025b; Li et al., 2025; Kang et al., 2026). These methods primarily use higher-order relations for representation learning, multimodal fusion, classification, or online adaptation (Li et al., 2026b), whereas target-conditioned, fixed-budget source-trial retrieval over higher-order source structures remains comparatively underexplored.

3 METHOD

This section is organized as follows. Section 3.1 formulates the target-conditioned source-trial retrieval problem. Section 3.2 describes source representation and construction of the source hypergraph. Section 3.3 presents target-conditioned source-trial retrieval and structural propagation. Section 3.4 details utility-supervised retriever learning and budgeted selection, and Section 3.5 describes adaptation using the selected source trials. An overview of the TETHER pipeline is presented in Figure 3.

3.1 PROBLEM SETTING

Let $\mathcal{D}_S = \{(\mathbf{X}_i, y_i, u_i)\}_{i=1}^{N_S}$ be the source pool, where $\mathbf{X}_i \in \mathbb{R}^{N_{\mathrm{ch}} \times T}$ is an EEG trial, $y_i \in \{1, \ldots, C\}$ is its label, and u_i is its source domain. A domain is a subject in within-dataset evaluation and a dataset–subject pair in cross-dataset evaluation. For an unseen target user, k labeled trials per class form the support set $\mathcal{S}_T$, while a disjoint query set $\mathcal{Q}_T$ is reserved for final evaluation.

Given a source budget B, class c receives $K_c = \lfloor B/C \rfloor + \mathbb{I}[c \le B \bmod C]$ trials, so that $\sum_c K_c = B$. The retriever selects $\mathcal{R}_T^c \subseteq \mathcal{D}_S^c$ with $|\mathcal{R}_T^c| = K_c$, and $\mathcal{R}_T = \bigcup_c \mathcal{R}_T^c$. We assume $K_c \le |\mathcal{D}_S^c|$. Source representations, the source hypergraph, and retriever parameters are learned from $\mathcal{D}_S$; $\mathcal{S}_T$ may condition retrieval and downstream decoder fitting, whereas $\mathcal{Q}_T$ is accessed only after the full pipeline is frozen.

3.2 SOURCE REPRESENTATION AND HYPERGRAPH CONSTRUCTION

Source-fitted representation. To capture complementary spatial and spectral information, each trial is represented by spatial covariance and spectral bandpower. The Riemannian reference and all feature standardizers are fitted on the source pool. The resulting multi-view feature is

$$\mathbf{f}_i^{\mathrm{ret}} = \left[\sqrt{\eta}\, \mathrm{Std}_S^{\tan}\big(\mathrm{TS}_S(\mathbf{C}_i)\big); \sqrt{1-\eta}\, \mathrm{Std}_S^{\mathrm{bp}}(\mathbf{b}_i) \right], \tag{1}$$

where $\mathbf{C}_i$ is the trial covariance, $\mathbf{b}_i$ is its channel-wise log-bandpower vector, TS_S is the tangent map at the source Riemannian mean (Barachant et al., 2012), $\mathrm{Std}_S^{\tan}$ and $\mathrm{Std}_S^{\mathrm{bp}}$ are source-fitted standardizers, and $\eta \in [0, 1]$ controls the two-view scaling.

A two-layer encoder ϕ_θ and a temporary classification head are trained on source labels. We then discard the head, freeze the encoder, and define the unit-normalized embedding $\mathbf{z}_i = \mathrm{norm}(\phi_{\theta^\star}(\mathbf{f}_i^{\mathrm{ret}})) \in \mathbb{R}^{d_{\mathrm{emb}}}$. The same frozen source-fitted transform embeds target support trials in the source embedding space.

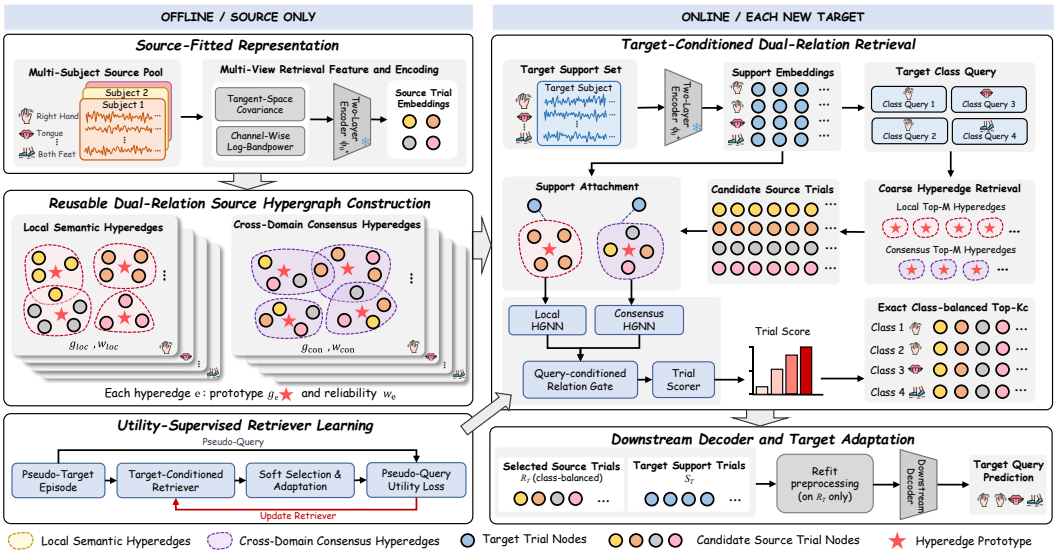


Figure 3: TETHER combines reusable source hypergraph construction, utility-supervised retriever learning, and target-conditioned source-trial retrieval for downstream adaptation.

Reusable dual-relation source hypergraph. For each class, we construct local semantic hyperedges for fine-grained neighborhoods and cross-domain consensus hyperedges for patterns recurring across source domains. Let $\mathcal{V}^c = \{v_i : y_i = c\}$, $\mathcal{V}_u = \{v_i : u_i = u\}$, and let $\mathcal{N}_i^c$ contain at most k_{pool} same-class cosine-nearest neighbors of anchor v_i. The local semantic hyperedge is

$$e_i^{\text{loc}} = \{v_i\} \cup \underset{v_j \in \mathcal{N}_i^c}{\arg\operatorname{top}^{k_{\text{loc}}}} \mathbf{z}_i^\top \mathbf{z}_j, \tag{2}$$

where k_{loc} is the number of retained neighbors.

Because local semantic neighborhoods may still reflect domain-specific structure, the cross-domain consensus hyperedge requires similar patterns to recur across source domains. The anchor is retained, while additional members are selected from other source domains using a constrained greedy procedure:

$$e_i^{\text{con}} = \{v_i\} \cup \text{GreedySelect}_{k_{\text{con}}} \left(\{ (v_j, \mathbf{z}_i^\top \mathbf{z}_j) : v_j \in \mathcal{N}_i^c \setminus \mathcal{V}_{u_i} \} \right), \tag{3}$$

where k_{con} is the neighbor count and GreedySelect covers at least $U_{\min}$ domains, including the anchor domain, while accepting at most b neighbors from each non-anchor domain. Infeasible cross-domain consensus hyperedges are discarded.

For relation $r \in \{\text{loc}, \text{con}\}$, the incidence matrix is $H_r^c(i, e) = \mathbb{I}[v_i \in e]$. Hyperedge e has prototype $\mathbf{g}_e = \text{norm}(|e|^{-1} \sum_{v_i \in e} \mathbf{z}_i)$ and reliability $w_e = |e|^{-1} \sum_{v_i \in e} (1 + \mathbf{z}_i^\top \mathbf{g}_e)/2$. The source hypergraph contains source trials only. The prototype supports target–hyperedge matching, while the reliability weight discounts weakly coherent hyperedges.

3.3 Target-Conditioned Dual-Relation Retrieval

Let $\mathcal{S}$ denote pseudo-support during retriever training or target support at deployment. Its class-c query is

$$\mathbf{q}_c(\mathcal{S}) = \text{norm}\left(\frac{1}{|\mathcal{S}^c|} \sum_{(\mathbf{X}_t, c) \in \mathcal{S}^c} \text{norm}\big(\phi_{\theta^\star}(\mathbf{f}_t^{\text{ret}})\big) \right), \tag{4}$$

where $\mathcal{S}^c$ contains the support trials labeled as class c.

Coarse hyperedge retrieval. We first restrict subsequent propagation to target-relevant source structure by retrieving

$$\widehat{\mathcal{E}}_r^c = \underset{e\in\mathcal{E}_r^c}{\overset{M}{\arg\operatorname{top}}} \left(w_e \, \mathbf{q}_c(\mathcal{S})^\top \mathbf{g}_e \right), \tag{5}$$

where $\mathcal{E}_r^c$ is the class-c hyperedge set for relation r, and M is the initial number retrieved per relation. The union of their source members forms $\mathcal{V}_{\text{cand}}^c$. If it contains fewer than K_c distinct trials, M is increased deterministically until the quota becomes feasible or all hyperedges are exhausted.

Support attachment and propagation. Coarse retrieval identifies relevant source patterns but does not yet inject target evidence into candidate representations. We therefore temporarily attach support node t to its m_{att} most similar retrieved hyperedges using

$$B_r^c(t,e) = \frac{\mathbb{I}[e \in \mathcal{N}_r(t)] \exp(\mathbf{z}_t^\top \mathbf{g}_e/\tau_b)}{\sum_{e'\in\mathcal{N}_r(t)} \exp(\mathbf{z}_t^\top \mathbf{g}_{e'}/\tau_b)}, \tag{6}$$

where $\mathcal{N}_r(t)$ is the attachment set and τ_b is the softmax temperature. Appending $\mathbf{B}_r^c$ to the retrieved source incidence matrix gives $\widetilde{\mathbf{H}}_r^c$. The normalized propagation operator is

$$\mathbf{P}_r^c = (\mathbf{D}_{v,r}^c)^{-1/2} \widetilde{\mathbf{H}}_r^c \widehat{\mathbf{\Omega}}_r^c (\mathbf{D}_{e,r}^c)^{-1} (\widetilde{\mathbf{H}}_r^c)^\top (\mathbf{D}_{v,r}^c)^{-1/2}, \tag{7}$$

where $\widehat{\mathbf{\Omega}}_r^c$ contains the retrieved-edge reliabilities, $\mathbf{D}_{e,r}^c = \operatorname{diag}((\widetilde{\mathbf{H}}_r^c)^\top \mathbf{1})$, and $\mathbf{D}_{v,r}^c = \operatorname{diag}(\widetilde{\mathbf{H}}_r^c \widehat{\mathbf{\Omega}}_r^c \mathbf{1})$. A zero inverse degree is assigned to any node absent from one relation branch. Separate two-layer hypergraph neural networks (HGNNs) (Feng et al., 2019) apply $\mathbf{P}_r^c$ to the stacked source-candidate and support embeddings, producing states $\mathbf{Z}_r^{(L),c}$.

Relation fusion and trial scoring. Because the local semantic and cross-domain consensus relations may contribute differently across targets and classes, a query-conditioned two-way softmax gate fuses the two branch states as $\overline{\mathbf{Z}}^c = \beta_{\text{loc}}^c \mathbf{Z}_{\text{loc}}^{(L),c} + \beta_{\text{con}}^c \mathbf{Z}_{\text{con}}^{(L),c}$. Candidate i is scored by

$$s_i^c = g_s\left([\overline{\mathbf{z}}_i^c; \mathbf{q}_c(\mathcal{S}); \overline{\mathbf{z}}_i^c \odot \mathbf{q}_c(\mathcal{S}); |\overline{\mathbf{z}}_i^c - \mathbf{q}_c(\mathcal{S})|; (\overline{\mathbf{z}}_i^c)^\top \mathbf{q}_c(\mathcal{S})] \right), \tag{8}$$

where $\overline{\mathbf{z}}_i^c$ is the fused state of candidate i, $\odot$ denotes element-wise multiplication, and g_s is a learnable scalar scorer.

3.4 Utility-Supervised Retriever Learning and Budgeted Selection

Pseudo-target optimization. To learn trial scores from downstream adaptation utility without accessing real target queries, we train the retriever through source-only pseudo-target episodes. Each episode holds out one source domain before constructing an episode-specific dual-relation source hypergraph and candidate pool from the remaining source domains. From the held-out source domain, k_ε per class form pseudo-support $\mathcal{S}_\varepsilon$ and the remainder form pseudo-query $\mathcal{Q}_\varepsilon$. The held-out source domain may have contributed to encoder pretraining, but never to the episode graph or source candidates. Retrieval is conditioned only on $\mathcal{S}_\varepsilon$; $\mathcal{Q}_\varepsilon$ is used solely for the utility loss.

Because exact Top-K_c selection is nondifferentiable, we replace it during training with budget-normalized soft weights:

$$\alpha_i^c = K_c \frac{\exp(s_i^c/\tau_s)}{\sum_{v_j\in\mathcal{V}_{\text{cand}}^c} \exp(s_j^c/\tau_s)}, \tag{9}$$

where τ_s is the weighting temperature. The nonnegative weights sum to K_c.

The soft weights enable differentiation but do not directly measure downstream adaptation utility. We therefore evaluate the weighted subset with a differentiable ridge estimator:

$$\mathbf{W}_\varepsilon = \left(\widetilde{\mathbf{F}}_\varepsilon^\top \mathbf{\Lambda}_\varepsilon \widetilde{\mathbf{F}}_\varepsilon + \lambda_{\text{ridge}} \mathbf{I}_0 \right)^{-1} \widetilde{\mathbf{F}}_\varepsilon^\top \mathbf{\Lambda}_\varepsilon \mathbf{Y}_\varepsilon, \tag{10}$$

where $\widetilde{\mathbf{F}}_\varepsilon$ contains frozen candidate and pseudo-support embeddings with an intercept column, $\mathbf{Y}_\varepsilon$ contains their one-hot labels, $\mathbf{\Lambda}_\varepsilon$ contains α_i^c for candidates and fixed weight γ for pseudo-support, λ_{ridge} is the ridge coefficient, and $\mathbf{I}_0$ leaves the intercept unregularized. The retriever minimizes

$$\psi^\star = \arg\min_\psi \mathbb{E}_\varepsilon \left[\operatorname{CE}\left(\widetilde{\mathbf{F}}_{\mathcal{Q}_\varepsilon} \mathbf{W}_\varepsilon, \mathbf{Y}_{\mathcal{Q}_\varepsilon} \right) \right], \tag{11}$$

where CE denotes the cross-entropy loss, ψ comprises the HGNN, gate, and scorer parameters, and $\widetilde{\mathbf{F}}_{\mathcal{Q}_\varepsilon}$ and $\mathbf{Y}_{\mathcal{Q}_\varepsilon}$ are the pseudo-query embeddings and labels. Gradients pass through the ridge solution (Bertinetto et al., 2019) and candidate weights; the encoder and discrete coarse-retrieval indices remain fixed.

Online exact selection. At deployment, all retriever components are frozen. The retriever uses the target support set $\mathcal{S}_T$ to retrieve and score source candidates, and the K_c highest-scoring distinct trials are selected from each class with deterministic tie breaking. Their union contains exactly B source trials.

3.5 Downstream Decoder

To prevent discarded source trials from influencing adaptation through preprocessing statistics, the Riemannian reference and both feature standardizers are refitted using $\mathcal{R}_T$ only. The frozen map transforms R_T and S_T, and a downstream decoder is fitted jointly on them. Selected source trials have unit weight; each target-support trial receives

$$w_T = \frac{\rho}{1-\rho}\frac{|\mathcal{R}_T|}{|\mathcal{S}_T|}, \tag{12}$$

where $\rho \in (0,1)$ is the target-support fraction of the total training weight. Thus, this fraction is invariant to the source budget and shot count. When a development decision is required, it uses an inner split of $\mathcal{S}_T$; the selected-source map and decoder are then refitted on the complete support set. The query set $\mathcal{Q}_T$ is transformed and evaluated once, after all statistics, parameters, and hyperparameters have been frozen.

4 Experiments

4.1 Experimental Setup

Datasets and protocols. We evaluate TETHER on three publicly available MI EEG datasets: BNCI2014-001 (Tangermann et al., 2012), Cho2017 (Cho et al., 2017), and Lee2019-MI (Lee et al., 2019). Four-class BNCI2014-001 is evaluated under within-dataset leave-one-subject-out (LOSO), whereas its left- and right-hand subset (BNCI-LR), Cho2017, and Lee2019-MI use cross-dataset leave-one-dataset-out (LODO), with the two non-target datasets forming the source pool. For each target subject, we sample $k \in \{5, 10, 20\}$ labeled calibration trials per class from a fixed calibration pool, with nested support sets $\mathcal{S}_5 \subset \mathcal{S}_{10} \subset \mathcal{S}_{20}$, and reserve a disjoint query set for final evaluation. The four-class task uses source budgets $B \in \{500, 1000\}$, while all binary tasks use $B = 500$; every fixed-budget method selects exactly B/C source trials per class. The target query set is never used for source ranking, source selection, feature-statistic estimation, model fitting, or hyperparameter selection. Dataset statistics, data partitions, and source-pool construction are detailed in Appendix A, while preprocessing and implementation details are provided in Appendix D.

Baselines and fairness. We compare TETHER with class-balanced random sampling, subject-level screening, pointwise spectral and Riemannian selection, pairwise k-nearest-neighbor (kNN) graph diffusion, and transfer-aware source-selection methods, including MTS-Top-B, MEIS-Top-B, and RSSDA-Top-B. All fixed-budget methods use identical target support sets, source pools, source budgets, and signal preprocessing. Subject-level Top-k + All Trials and All-source are treated as unrestricted references and excluded from the fixed-budget ranking. Baseline definitions and their fixed-budget adaptations are provided in Appendix C.

Evaluation and statistics. Each shot-dependent configuration is repeated with five shared random seeds. Results are first averaged over seeds within each target subject and then summarized across subjects as mean $\pm$ sample standard deviation. Macro Balanced Accuracy (Macro BA) is the primary evaluation metric, with per-class recall reported for complementary analysis. The target subject, rather than the random seed, is treated as the paired statistical unit. For each of the three binary LODO tasks, we apply a Friedman omnibus test within each shot setting, followed by two-sided paired Wilcoxon signed-rank tests comparing TETHER with each baseline and Holm correction for multiple comparisons. Metric definitions and statistical procedures are given in Appendix B.

4.2 MAIN RESULTS

The main results assess three aspects of TETHER: source-trial retrieval effectiveness under within-dataset subject shift, generalization under cross-dataset shift, and the computational efficiency enabled by reusable source organization.

Cross-subject generalization (LOSO). Tables 1 and 2 evaluate whether target-conditioned source-trial retrieval with higher-order structure improves few-shot adaptation over subject-level, pointwise, pairwise, and transfer-aware source selection under the same source budget. TETHER achieves the highest macro BA among all fixed-budget methods across all six budget–shot settings, outperforming the strongest competing method by more than 10 percentage points in every setting. TETHER also maintains more balanced class-wise performance, supporting the benefit of combining the local semantic and cross-domain consensus relations with utility-supervised retriever learning. Moreover, doubling the source budget from 500 to 1000 yields only marginal additional gains, indicating that effective adaptation depends more on selecting high-utility trials than on simply retaining more source data.

Table 1: Within-dataset cross-subject performance on four-class BNCI2014-001 under LOSO evaluation with $B = 500$. Bold and underlined values denote the best and second-best fixed-budget results, respectively; unrestricted references are excluded from ranking.

Shot	Method	Per-class Recall ↑				Macro BA ↑
		Left Hand	Right Hand	Both Feet	Tongue	
5-shot	Class-balanced Random-B	0.4355±0.1076	0.2472±0.0752	0.2056±0.0895	0.2457±0.1004	0.2835±0.0366
	Subject-level Top-1 + Random-B	0.3954±0.2095	0.2364±0.1017	0.1806±0.0920	0.2278±0.0852	0.2600±0.0165
	Subject-level Top-1 + All Trials	0.4722±0.1836	0.2981±0.0818	0.2105±0.1220	0.1935±0.1157	0.2936±0.0645
	Bandpower Top-B	0.3948±0.1868	0.2926±0.1152	0.1907±0.1211	0.2670±0.1210	0.2863±0.0563
	Riemannian Top-B	0.2731±0.2156	0.3253±0.2540	0.1111±0.1327	<u>0.3985</u>±0.3028	0.2770±0.0424
	Pairwise kNN Graph Diffusion	0.3475±0.1150	0.3676±0.1763	0.2182±0.1164	0.3611±0.2178	0.3236±0.0730
	MTS-Top-B (Liang et al., 2021)	0.4377±0.1461	<u>0.5627</u>±0.1962	0.1210±0.1358	0.3605±0.3684	<u>0.3704</u>±0.1285
	MEIS-Top-B (Liang et al., 2024)	**0.4852**±0.2635	0.4000±0.3269	0.1630±0.1204	0.2164±0.1543	0.3161±0.0863
	RSSDA-Top-B (Ding et al., 2026)	0.3296±0.1348	0.3262±0.1268	<u>0.2858</u>±0.0957	0.3793±0.1442	0.3302±0.0837
	TETHER (Ours)	<u>0.4830</u>±0.1417	**0.5654**±0.1610	**0.4691**±0.1358	**0.5608**±0.2231	**0.5196**±0.1425
10-shot	Class-balanced Random-B	0.4130±0.1230	0.3414±0.1464	0.2037±0.0769	0.2364±0.1140	0.2986±0.0511
	Subject-level Top-1 + Random-B	0.3907±0.2126	0.2904±0.1331	0.1778±0.0471	0.2219±0.0764	0.2702±0.0309
	Subject-level Top-1 + All Trials	0.4556±0.1899	0.3849±0.0871	0.1497±0.0767	0.2969±0.2027	0.3218±0.0816
	Bandpower Top-B	0.3824±0.1901	0.3522±0.1268	0.1966±0.0609	0.3287±0.1183	0.3150±0.0590
	Riemannian Top-B	0.2985±0.2430	0.3676±0.2463	0.1031±0.0734	<u>0.4340</u>±0.2427	0.3008±0.0538
	Pairwise kNN Graph Diffusion	0.4420±0.1888	0.3398±0.1652	0.1744±0.0475	0.2799±0.0914	0.3090±0.0558
	MTS-Top-B (Liang et al., 2021)	0.4244±0.2244	**0.6460**±0.1586	0.1003±0.0951	0.2244±0.2725	<u>0.3488</u>±0.1194
	MEIS-Top-B (Liang et al., 2024)	<u>0.5392</u>±0.2459	0.4219±0.3046	0.1182±0.0863	0.2247±0.1610	0.3260±0.1018
	RSSDA-Top-B (Ding et al., 2026)	0.3691±0.1287	0.3472±0.1231	<u>0.2963</u>±0.1260	0.3546±0.1185	0.3418±0.0905
	TETHER (Ours)	**0.5457**±0.1892	<u>0.5627</u>±0.1960	**0.4849**±0.1558	**0.5694**±0.2433	**0.5407**±0.1440
20-shot	Class-balanced Random-B	0.4330±0.1041	0.3694±0.1859	0.2315±0.1030	0.3623±0.2025	0.3491±0.1131
	Subject-level Top-1 + Random-B	0.3917±0.1734	0.3586±0.1464	0.2219±0.0735	0.3028±0.1354	0.3188±0.0883
	Subject-level Top-1 + All Trials	0.4343±0.2339	0.4231±0.1862	0.2015±0.1035	0.3281±0.1529	0.3468±0.0864
	Bandpower Top-B	0.3475±0.1481	0.4407±0.2298	0.2404±0.0940	0.4164±0.2720	0.3613±0.1137
	Riemannian Top-B	0.3142±0.2894	0.3985±0.2992	0.1204±0.1025	<u>0.4688</u>±0.2351	0.3255±0.0917
	Pairwise kNN Graph Diffusion	0.3985±0.1893	0.4031±0.1970	0.2398±0.0997	0.3951±0.1858	0.3591±0.1179
	MTS-Top-B (Liang et al., 2021)	<u>0.5090</u>±0.1633	<u>0.5256</u>±0.2078	0.1753±0.1168	0.3651±0.2845	<u>0.3938</u>±0.1382
	MEIS-Top-B (Liang et al., 2024)	0.4262±0.1962	0.4466±0.2439	0.1864±0.0889	0.2728±0.1847	0.3330±0.0974
	RSSDA-Top-B (Ding et al., 2026)	0.3139±0.1649	0.3469±0.1528	<u>0.2716</u>±0.0660	0.4485±0.1886	0.3452±0.0941
	TETHER (Ours)	**0.5485**±0.2103	**0.5722**±0.2319	**0.5673**±0.1898	**0.6244**±0.2292	**0.5781**±0.1301
–	All-source	0.6059±0.1721	0.5861±0.2146	0.3562±0.2352	0.3778±0.2816	0.4815±0.1729

Cross-dataset generalization (LODO). Figure 4 examines whether the learned retrieval strategy remains effective when the target dataset is excluded from the source pool. TETHER achieves the highest mean macro BA among the fixed-budget methods across all three target datasets and all three target-shot settings. Its consistent advantage across datasets with different subject populations and recording configurations indicates that the learned retrieval strategy extends beyond within-dataset subject matching to heterogeneous external-source settings. Complete numerical results for the binary cross-dataset experiments, including per-class recall and macro F1, are provided in Appendix E.

Computational efficiency. TETHER also substantially reduces per-target adaptation cost. Its online source-trial retrieval and target adaptation require only 0.641–0.647 s per condition, approximately $9\times$ faster than the fastest competing fixed-budget method, while the amortized full-pipeline runtime remains below all compared methods. This supports the organize-once, retrieve-per-target design, in which reusable source-side computation is performed offline and only lightweight target-conditioned retrieval is required for each new user. Complete runtime results are reported in Appendix E.

Table 2: Within-dataset cross-subject performance on four-class BNCI2014-001 under LOSO evaluation with $B = 1000$. Bold and underlined values denote the best and second-best fixed-budget results, respectively; unrestricted references are excluded from ranking.

Shot	Method	Per-class Recall ↑				Macro BA ↑
		Left Hand	Right Hand	Both Feet	Tongue	
5-shot	Class-balanced Random-B	0.4568±0.1306	0.4599±0.1421	0.2120±0.0994	0.4025±0.3079	0.3828±0.1040
	Subject-level Top-2 + Random-B	0.4707±0.1765	0.4019±0.2137	0.2414±0.1726	0.3438±0.2783	0.3644±0.1051
	Subject-level Top-2 + All Trials	0.5228±0.1588	0.3769±0.1830	0.1855±0.1093	0.3778±0.2927	0.3657±0.1059
	Bandpower Top-B	0.3478±0.1665	0.5361±0.2493	0.2423±0.1638	0.4370±0.3228	0.3908±0.1193
	Riemannian Top-B	0.3923±0.1530	0.4068±0.1709	0.2444±0.1145	0.4809±0.2697	0.3811±0.1145
	Pairwise kNN Graph Diffusion	0.4034±0.1751	0.4485±0.1351	0.2586±0.1316	0.4713±0.2723	0.3954±0.1151
	MTS-Top-B (Liang et al., 2021)	0.4861±0.1918	**0.5815**±0.2014	0.2009±0.1804	0.4130±0.3249	0.4204±0.1465
	MEIS-Top-B (Liang et al., 2024)	0.5262±0.1666	0.4414±0.1818	0.2525±0.1193	0.4395±0.2704	0.4149±0.1509
	RSSDA-Top-B (Ding et al., 2026)	0.3500±0.1421	0.3602±0.1474	0.2812±0.0878	0.4228±0.1570	0.3535±0.0938
	TETHER (Ours)	**0.5287**±0.1997	0.5127±0.2101	**0.4747**±0.1568	**0.5769**±0.2218	**0.5233**±0.1344
10-shot	Class-balanced Random-B	0.5164±0.1115	0.4738±0.1041	0.2475±0.1332	0.3744±0.2636	0.4030±0.1188
	Subject-level Top-2 + Random-B	0.4528±0.1496	0.4759±0.1653	0.2398±0.1466	0.3778±0.2776	0.3866±0.1217
	Subject-level Top-2 + All Trials	0.5284±0.1570	0.4170±0.1825	0.2651±0.1670	0.3728±0.3065	0.3958±0.1364
	Bandpower Top-B	0.3901±0.2142	**0.6086**±0.2282	0.2148±0.1260	0.4843±0.3590	0.4245±0.1383
	Riemannian Top-B	0.4059±0.1810	0.4336±0.1471	0.2722±0.1371	0.5093±0.2632	0.4052±0.1338
	Pairwise kNN Graph Diffusion	0.3670±0.1867	0.5336±0.1925	0.3145±0.1861	0.4488±0.2988	0.4160±0.1325
	MTS-Top-B (Liang et al., 2021)	0.5006±0.1651	0.5497±0.1606	0.2201±0.1772	0.4497±0.3271	0.4300±0.1503
	MEIS-Top-B (Liang et al., 2024)	**0.5540**±0.2033	0.4917±0.1925	0.2296±0.0909	0.4012±0.2674	0.4191±0.1490
	RSSDA-Top-B (Ding et al., 2026)	0.4034±0.1550	0.3694±0.1503	0.2432±0.0825	0.4046±0.1933	0.3552±0.0972
	TETHER (Ours)	0.5538±0.2084	0.5464±0.2337	**0.5190**±0.1976	**0.6112**±0.2515	**0.5576**±0.1373
20-shot	Class-balanced Random-B	0.5744±0.1186	0.4639±0.1829	0.2494±0.1371	0.4157±0.2765	0.4258±0.1420
	Subject-level Top-2 + Random-B	0.5068±0.1545	0.5102±0.2292	0.2216±0.1539	0.3802±0.2962	0.4047±0.1437
	Subject-level Top-2 + All Trials	0.5333±0.1771	0.4932±0.2657	0.2262±0.1561	0.4157±0.3235	0.4171±0.1437
	Bandpower Top-B	0.4485±0.2084	0.5778±0.2280	0.2571±0.1361	0.5247±0.3238	0.4520±0.1506
	Riemannian Top-B	0.4451±0.1966	0.4935±0.2247	0.2852±0.1300	0.5444±0.2369	0.4421±0.1630
	Pairwise kNN Graph Diffusion	0.4670±0.1518	0.5170±0.1407	0.2966±0.1788	0.4546±0.2903	0.4338±0.1438
	MTS-Top-B (Liang et al., 2021)	**0.5836**±0.1424	**0.5914**±0.1697	0.2410±0.1591	0.4602±0.3540	0.4691±0.1701
	MEIS-Top-B (Liang et al., 2024)	0.5790±0.1853	0.4713±0.2231	0.2463±0.1152	0.4846±0.3260	0.4453±0.1831
	RSSDA-Top-B (Ding et al., 2026)	0.3738±0.1773	0.3920±0.1380	0.2596±0.0701	0.4688±0.2226	0.3735±0.1165
	TETHER (Ours)	0.5774±0.2036	0.5715±0.2271	**0.5752**±0.2338	**0.6873**±0.1970	**0.6029**±0.1393
–	All-source	0.6059±0.1721	0.5861±0.2146	0.3562±0.2352	0.3778±0.2816	0.4815±0.1729

4.3 Ablation Study

We conduct ablation studies to assess the contributions of the dual-relation source hypergraph, structural propagation, and utility-supervised retriever learning. Table 3 compares four variants on BNCI2014-001 with $B = 500$: removing the cross-domain consensus relation, removing the local semantic relation, disabling HGNN propagation, and removing utility supervision together with the learned trial scorer.

The complete model achieves the highest macro BA across all three target-shot settings, although improvements are not uniform across individual classes. Removing utility supervision produces the largest and most consistent degradation, indicating that higher-order structure alone is insufficient to determine which structurally relevant source trials are genuinely useful for target adaptation. Removing either hypergraph relation or HGNN propagation also degrades performance, while the relatively similar behavior of these structural ablations suggests that the local semantic relation, the cross-domain consensus relation, and relational propagation provide complementary rather than interchangeable information. Overall, the ablation results support a coordinated design in which the source hypergraph organizes transferable source patterns, structural propagation contextualizes



Figure 4: Mean macro BA, subject-level standard deviation, and Holm-adjusted p-values from Wilcoxon signed-rank tests comparing each competing method with TETHER under LODO ($B = 500$). *: $0.01 \leq p_{\text{Holm}} < 0.05$, **: $0.001 \leq p_{\text{Holm}} < 0.01$, and ***: $p_{\text{Holm}} < 0.001$.

them with respect to the target, and utility supervision aligns structural relevance with downstream adaptation benefit.

Table 3: Ablation results on BNCI2014-001 with $B = 500$. Bold and underlined values denote the best and second-best results.

Shot	Method	Per-class Recall ↑				Macro BA ↑
		Left Hand	Right Hand	Both Feet	Tongue	
5-shot	w/o Cross-Domain Consensus Relation	0.4491±0.1516	0.5182±0.1978	0.4571±0.1584	0.5441±0.2603	0.4921±0.1338
	w/o Local Semantic Relation	0.4602±0.1844	0.5160±0.1849	0.4370±0.1141	**0.5608**±0.2540	0.4935±0.1379
	w/o HGNN Propagation	0.4608±0.1795	0.5077±0.2112	0.4552±0.1691	0.5556±0.2326	0.4948±0.1310
	w/o Utility Supervision	0.4642±0.1701	0.4806±0.1951	0.4454±0.1782	0.5463±0.2321	0.4841±0.1277
	TETHER (Ours)	**0.4830**±0.1417	**0.5654**±0.1610	**0.4691**±0.1358	**0.5608**±0.2231	**0.5196**±0.1425
10-shot	w/o Cross-Domain Consensus Relation	0.4818±0.1981	0.5401±0.2224	0.4867±0.2099	0.5840±0.2434	0.5231±0.1437
	w/o Local Semantic Relation	0.4948±0.1846	0.5426±0.2337	0.4877±0.1648	0.5806±0.2654	0.5264±0.1525
	w/o HGNN Propagation	0.4892±0.1815	0.5343±0.2384	**0.4926**±0.2052	0.5904±0.2367	0.5266±0.1386
	w/o Utility Supervision	0.4806±0.1793	0.5262±0.2490	0.4849±0.2058	**0.5907**±0.2442	0.5206±0.1379
	TETHER (Ours)	**0.5457**±0.1892	**0.5627**±0.1960	0.4849±0.1558	0.5694±0.2433	**0.5407**±0.1440
20-shot	w/o Cross-Domain Consensus Relation	0.5068±0.1982	0.5673±0.2687	0.5611±0.2190	0.6281±0.2301	0.5658±0.1407
	w/o Local Semantic Relation	0.5145±0.2112	0.5562±0.2504	0.5549±0.1730	0.6275±0.2358	0.5633±0.1527
	w/o HGNN Propagation	0.5139±0.1962	0.5537±0.2728	0.5438±0.2112	**0.6358**±0.2342	0.5618±0.1547
	w/o Utility Supervision	0.4978±0.1909	0.5694±0.2571	0.5466±0.2016	0.6154±0.2613	0.5573±0.1395
	TETHER (Ours)	**0.5485**±0.2103	**0.5722**±0.2319	**0.5673**±0.1898	0.6244±0.2292	**0.5781**±0.1301

5 Conclusion

In this work, we revisit few-shot MI EEG adaptation by casting it as target-conditioned source-trial retrieval, where transferability is viewed as a target-dependent structural utility rather than an intrinsic property of individual trials. This perspective motivates TETHER, an organize-once, retrieve-per-target framework that organizes historical EEG into a reusable dual-relation source hypergraph and performs utility-supervised retriever learning through source-only pseudo-target episodes. Across within-dataset cross-subject and cross-dataset evaluations, TETHER consistently improves fixed-budget adaptation while enabling efficient reuse of the source pool. The primary contribution is therefore not only the specific hypergraph retriever, but also a retrieval perspective that identifies target-relevant source structures according to their downstream adaptation utility. Our empirical conclusions are currently limited to MI EEG. The proposed retrieval design assumes labeled source domains, a shared label space between source and target data, and representations that support meaningful cross-domain neighborhood comparisons. Extending and validating this design beyond MI EEG, including other EEG paradigms and sensing modalities, remains an important direction for future work.

AI USE STATEMENT

We used generative AI tools to assist with language polishing of the manuscript and with the retrieval and discovery of relevant literature. All AI-assisted edits and candidate references were manually reviewed and verified by the authors. We take full responsibility for the final content of this work.

REPRODUCIBILITY

We provide code for our method and implementations of all evaluated baselines in the supplementary package TETHER_code, together with detailed usage instructions and configuration files. Hyperparameters and evaluation protocols are described in Appendix D.

REFERENCES

Sion An, Myeongkyun Kang, Soopil Kim, Philip Chikontwe, Li Shen, and Sang Hyun Park. Subject-adaptive transfer learning using resting state EEG signals for cross-subject EEG motor imagery classification. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume 15011 of *Lecture Notes in Computer Science*, pp. 678–688. Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-72120-5_63.

Alessia Antelmi, Gennaro Cordasco, Mirko Polato, Vittorio Scarano, Carmine Spagnuolo, and Dingqi Yang. A survey on hypergraph representation learning. *ACM Computing Surveys*, 56(1): 24:1–24:38, 2023. doi: 10.1145/3605776. URL https://dl.acm.org/doi/10.1145/3605776.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012. doi: 10.1109/TBME.2011.2172210.

Luca Bertinetto, João F. Henriques, Philip H. S. Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *International Conference on Learning Representations*, 2019.

Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010. doi: 10.1109/TSP.2010.2053029.

Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are AllSet: A multiset function framework for hypergraph neural networks. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=hpBTIv2uy_E.

Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. EEG datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 2017. doi: 10.1093/gigascience/gix034.

Arnav M. Das, Gantavya Bhatt, Sahil Verma, Yiping Wang, Viswa Virinchi Muppirala, and Jeff Bilmes. Matched data, better models: Target aligned data filtering with sparse autoencoders. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8291acb610ddc260d554da5a3d273ae9-Abstract-Conference.html.

Tao Ding, Yuhao Zhang, Xingwei Zhao, Jian Shi, Jun Liu, Zhouping Tang, Bo Tao, and Han Ding. RSSDA: A Riemannian source selection and domain alignment framework for cross-subject multi-class motor imagery classification. *Pattern Recognition*, 180:113934, 2026. doi: 10.1016/j.patcog.2026.113934.

Yi Ding, Neethu Robinson, Chengxuan Tong, Qiuhao Zeng, and Cuntai Guan. LGGNet: Learning from local-global-graph representations for brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7):9773–9786, 2024. doi: 10.1109/TNNLS.2023.3236635.

Yassine El Ouahidi, Vincent Gripon, Bastien Pasdeloup, Ghaith Bouallegue, Nicolas Farrugia, and Giulia Lioi. A strong and simple deep learning baseline for BCI motor imagery decoding. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:3338–3347, 2024. doi: 10.1109/TNSRE.2024.3451010.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558.

Chang Gao, Xiaohu Li, Sihui Cheng, Wenyuan Liu, and Siyao Guo. Hybrid curriculum and transfer learning method for cross-subject motor imagery recognition. *Expert Systems with Applications*, 299:130221, 2026. doi: 10.1016/j.eswa.2025.130221.

Daeun Gwon and Minkyu Ahn. Motor task-to-task transfer learning for motor imagery brain-computer interfaces. *NeuroImage*, 302:120906, 2024. doi: 10.1016/j.neuroimage.2024.120906.

Qiaomei Huang, Zikun Cai, and Tian-jian Luo. Multiple regularized knowledge transfer learning for cross-subject motor imagery electroencephalogram classification. *Engineering Applications of Artificial Intelligence*, 165:113397, 2026. doi: 10.1016/j.engappai.2025.113397.

Ayush Jain, Andrea Montanari, and Eren Sasoglu. Train on validation (ToV): Fast data selection with applications to fine-tuning. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/1c3d419b754cb4de0a67a453cb28d959-Abstract-Conference.html.

Wei-Bang Jiang, Li-Ming 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.

Jing Jin, Guanglian Bai, Ren Xu, Ke Qin, Hao Sun, Xingyu Wang, and Andrzej Cichocki. A cross-dataset adaptive domain selection transfer learning framework for motor imagery-based brain-computer interfaces. *Journal of Neural Engineering*, 21(3):036057, 2024. doi: 10.1088/1741-2552/ad593b.

Zijian Kang, Yueyang Li, Shengyu Gong, Weiming Zeng, Hongjie Yan, Lingbin Bian, Zhiguo Zhang, Wai Ting Siok, and Nizhuan Wang. Hypergraph multi-modal learning for EEG-based emotion recognition in conversation. *Neural Networks*, 202:109057, 2026. doi: 10.1016/j.neunet.2026.109057.

Krishnateja Killamsetty, Durga Sivasubramanian, Ganesh Ramakrishnan, Abir De, and Rishabh Iyer. GRAD-MATCH: Gradient matching based data subset selection for efficient deep model training. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5464–5474. PMLR, 2021a. URL https://proceedings.mlr.press/v139/killamsetty21a.html.

Krishnateja Killamsetty, Durga Sivasubramanian, Ganesh Ramakrishnan, and Rishabh Iyer. GLISTER: Generalization based data subset selection for efficient and robust learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 8110–8118, 2021b. doi: 10.1609/aaai.v35i9.16988. URL https://ojs.aaai.org/index.php/AAAI/article/view/16988.

Min-Ho Lee, O-Yeon Kwon, Yong-Jeong Kim, Hong-Kyung Kim, Young-Eun Lee, John Williamson, Siamac Fazli, and Seong-Whan Lee. EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*, 8(5):giz002, 2019. doi: 10.1093/gigascience/giz002.

Fan Li, Xiaoyang Wang, Wenjie Zhang, Ying Zhang, and Xuemin Lin. DHG-Bench: A comprehensive benchmark for deep hypergraph learning. In *International Conference on Learning Representations*, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/4f25c9511feb3d6496450d68ca21efb9-Abstract-Conference.html.

Siqi Li, Zhi Li, Tong Liu, Shuai Zhang, Yanfei Jia, Zhiqiang Yi, Jue Xie, and Ni Ji. Multi-feature Riemannian hypergraph for online test-time adaptation of motor imagery brain–computer interface. *arXiv preprint arXiv:2608.16134*, 2026b. doi: 10.48550/arXiv.2608.16134. URL https://arxiv.org/abs/2608.16134.

Yibing Li, Zhenye Zhao, Jiangchuan Liu, Yong Peng, Kenneth Camilleri, Wanzeng Kong, and Andrzej Cichocki. EEG-based speech imagery decoding by dynamic hypergraph learning within projected and selected feature subspaces. *Journal of Neural Engineering*, 22(4):046030, 2025. doi: 10.1088/1741-2552/adeec8.

Zilin Liang, Zheng Zheng, Weihai Chen, Jianhua Wang, Jianbin Zhang, Jianer Chen, and Zuobing Chen. Manifold trial selection to reduce negative transfer in motor imagery-based brain–computer interface. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4144–4149. IEEE, 2021. doi: 10.1109/IROS51168.2021.9636137.

Zilin Liang, Zheng Zheng, Weihai Chen, Zhongcai Pei, Jianhua Wang, and Jianer Chen. Manifold embedded instance selection to suppress negative transfer in motor imagery-based brain–computer interface. *Biomedical Signal Processing and Control*, 88:105556, 2024. doi: 10.1016/j.bspc.2023.105556.

Tian-jian Luo. Dual regularized spatial-temporal features adaptation for multi-source selected cross-subject motor imagery EEG classification. *Expert Systems with Applications*, 255:124673, 2024. doi: 10.1016/j.eswa.2024.124673.

Han Wei Ng and Cuntai Guan. Subject-independent meta-learning framework towards optimal training of EEG-based classifiers. *Neural Networks*, 172:106108, 2024. doi: 10.1016/j.neunet.2024.106108.

Kecheng Shi, Rui Huang, Jianzhi Lyu, Zhe Li, Fengjun Mu, Zhinan Peng, Chaobin Zou, Hong Cheng, Jianwei Zhang, and Bijoy Kumar Ghosh. EEG-based motor imagery classification with tuned heuristic fusion graph convolutional network for rehabilitation training. *IEEE Transactions on Automation Science and Engineering*, 22:14928–14939, 2025. doi: 10.1109/TASE.2025.3564162.

Hao Sun, Yi Ding, Jianzhu Bao, Ke Qin, Chengxuan Tong, Jing Jin, and Cuntai Guan. Leveraging temporal dependency for cross-subject-MI BCIs by contrastive learning and self-attention. *Neural Networks*, 178:106470, 2024. doi: 10.1016/j.neunet.2024.106470.

Bohan Tang, Zexi Liu, Keyue Jiang, Siheng Chen, and Xiaowen Dong. Training-free message passing for learning on hypergraphs. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=4AuyYxt7A2.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot R. Müller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6:55, 2012. doi: 10.3389/fnins.2012.00055.

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, pp. 39249–39280, 2024. doi: 10.52202/079017-1239. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4540d267eeec4e5dbd9dae9448f0b739-Abstract-Conference.html.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations*, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/bbbd6d915cb90be21c1254a82d45cedd-Abstract-Conference.html.

Peihao Wang, Shenghao Yang, Yunyu Liu, Zhangyang Wang, and Pan Li. Equivariant hypergraph diffusion neural operators. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=RiTjKoscnNd.

Pengpai Wang, Tiantian Xie, Yueying Zhou, Peiliang Gong, and Rosa H. M. Chan. TCPL: Task-conditioned prompt learning for few-shot cross-subject motor imagery EEG decoding. *Frontiers in Neuroscience*, 19:1689286, 2025b. doi: 10.3389/fnins.2025.1689286.

Yihan Wang, Jiaxing Wang, Weiqun Wang, Jianqiang Su, Chayut Bunterngchit, and Zeng-Guang Hou. TFTL: A task-free transfer learning strategy for EEG-based cross-subject and cross-dataset motor imagery BCI. *IEEE Transactions on Biomedical Engineering*, 72(2):810–821, 2025c. doi: 10.1109/TBME.2024.3474049.

Dongrui Wu. Revisiting Euclidean alignment for transfer learning in EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 22(3):031005, 2025. doi: 10.1088/1741-2552/addd49.

Jielong Yan, Yifan Feng, Shihui Ying, and Yue Gao. Hypergraph dynamic system. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/0973524e02a712af33325d0688ae6f49-Abstract-Conference.html.

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, 2023. doi: 10.52202/075280-3420. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/f6b30f3e2dd9cb53bbf2024402d02295-Abstract.html.

Siqi Yang, Zhihua Huang, and Tian-jian Luo. Joint spatial feature adaption and confident pseudo-label selection for cross-subject motor imagery EEG signals classification. *Expert Systems with Applications*, 278:127312, 2025. doi: 10.1016/j.eswa.2025.127312.

Junjie Oscar Yin and Alexander M. Rush. Compute-constrained data selection. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/5acb720a361eecb34ee62d356859d246-Abstract-Conference.html.

Kang Yin, Elissa Yanting Lim, and Seong-Whan Lee. GITGAN: Generative inter-subject transfer for EEG motor imagery analysis. *Pattern Recognition*, 146:110015, 2024. doi: 10.1016/j.patcog.2023.110015.

Fan Zhang, Hanliang Wu, and Yuxin Guo. Semi-supervised multi-source transfer learning for cross-subject EEG motor imagery classification. *Medical & Biological Engineering & Computing*, 62 (6):1655–1672, 2024. doi: 10.1007/s11517-024-03032-z.

Jiayang Zhang, Kang Li, Banghua Yang, and Zhengrun Zhao. Cross-dataset motor imagery decoding—a transfer learning assisted graph convolutional network approach. *Biomedical Signal Processing and Control*, 102:107213, 2025a. doi: 10.1016/j.bspc.2024.107213.

Yan Zhang, Yang Gao, and Min Li. Joint modeling of fMRI and EEG imaging using ordinary differential equation-based hypergraph neural networks. In *Advances in Neural Information Processing Systems*, volume 38, pp. 158753–158771, 2025b. doi: 10.52202/085713-4790.

Minmin Zheng and Yiwen Lin. A deep transfer learning network with two classifiers based on sample selection for motor imagery brain-computer interface. *Biomedical Signal Processing and Control*, 89:105786, 2024. doi: 10.1016/j.bspc.2023.105786.

Hongyi Zhi, Tianyou Yu, Zhenghui Gu, Zhuobin Lin, Le Che, Yuanqing Li, and Zhuliang Yu. Supervised contrastive learning-based domain generalization network for cross-subject motor decoding. *IEEE Transactions on Biomedical Engineering*, 72(1):401–412, 2025. doi: 10.1109/TBME.2024.3432934.

Fenfang Zhu, Jicheng Cai, Hao Zheng, Zilin Liang, and Yue Zhang. Suppression of negative transfer in motor imagery brain–computer interface based on mutual information and Pearson correlation coefficient. *Review of Scientific Instruments*, 95(7):075105, 2024. doi: 10.1063/5.0208524.

APPENDIX ARRANGEMENT

The Appendix is organized as follows:

- **Section A:** Datasets and data partitions.
- **Section B:** Evaluation metrics and statistical analysis.
- **Section C:** Baseline definitions and fixed-budget adaptations.
- **Section D:** Implementation details and hyperparameters.
- **Section E:** Complete binary cross-dataset results.

A DATASETS AND DATA PARTITIONING

We use three publicly available cue-based motor-imagery (MI) EEG datasets. Table 4 summarizes the scale and acquisition characteristics of their original MI recordings. Only EEG recordings from the MI paradigms are used in our experiments.

Table 4: Summary of the MI EEG datasets. Trial counts refer to the original MI recordings before constructing task-specific experimental subsets.

Dataset	# Subjects	# MI classes	# EEG ch.	Hz	Recording structure	# Trials/subject	# Total trials
BNCI2014-001	9	4 (LH/RH/F/T)	22	250	2 sessions $\times$ 6 runs	576	5,184
Cho2017	52	2 (LH/RH)	64	512	5 runs (6 for S07/S09/S46)	200/240	10,520
Lee2019-MI	54	2 (LH/RH)	62	1,000	2 sessions $\times$ 2 phases	400	21,600
Total	**115**	–	–	–	–	–	**37,304**

BNCI2014-001. BNCI2014-001, also known as BCI Competition IV Dataset 2a, is a widely used benchmark for four-class MI decoding and contains left-hand, right-hand, both-feet, and tongue imagery tasks (Tangermann et al., 2012). Each trial begins with a fixation cross and an acoustic warning, followed by a visual cue at $t = 2\,\mathrm{s}$; the subject then performs the indicated imagery task until $t = 6\,\mathrm{s}$. The recordings were collected on different days, introducing session-dependent variability in addition to the substantial differences among subjects. The original release also contains EOG measurements for artifact monitoring, but only the EEG channels are used in our experiments. We retain all four MI classes for the four-class evaluation. We additionally construct BNCI-LR by retaining only the left- and right-hand trials, initially yielding 288 trials per subject and 2,592 trials in total. In this binary preprocessing pipeline, we exclude trials containing an official GDF rejection marker within trial intervals defined by trial-start events. This removes 246 trials, leaving 2,346 trials: 1,163 from Session 1 and 1,183 from Session 2.

Cho2017. Cho2017 is a large-cohort binary MI dataset designed to support the analysis of inter-subject variability in EEG-based BCI systems (Cho et al., 2017). The experimental task consists of visually cued left- and right-hand motor imagery. Each trial comprises a 2-s fixation period followed by a 3-s instruction period, during which the subject performs the indicated imagery task. Consecutive trials are separated by a randomized resting interval to reduce temporal predictability and carry-over effects between trials. In addition to the task EEG, the original release provides resting-state and auxiliary physiological recordings, including EMG signals that can be used to monitor unintended muscle activity. We retain only the cue-locked EEG trials from the binary MI experiment, ensuring that every retained trial corresponds to one of the two target classes. Its broad subject population makes Cho2017 particularly suitable for evaluating whether a source-selection method can accommodate heterogeneous neural patterns across users.

Lee2019-MI. Lee2019-MI is the MI component of the OpenBMI benchmark and considers imagined left- and right-hand grasping (Lee et al., 2019). The recordings span different days and separate training and test phases, thereby capturing both inter-day variability and within-subject changes across experimental phases. Each trial consists of a 3-s fixation period followed by a directional cue

and 4 s of motor imagery. OpenBMI was designed as a multi-paradigm BCI resource and additionally provides ERP, SSVEP, and auxiliary physiological recordings. We use only the offline-training EEG recordings EEG_MI_train from both sessions, excluding EEG_MI_test, all non-MI paradigms, and auxiliary signals. This yields 200 trials per subject and 10,800 trials in total, with 5,400 trials from each session. The repeated recordings make Lee2019-MI a challenging benchmark for assessing target adaptation under subject- and session-dependent EEG nonstationarity.

Target calibration and query partitions. Target calibration pools and query sets are fixed before shot sampling and are shared across all compared methods. For each target subject and random seed, 5, 10, and 20 labeled trials per class are sampled from the calibration pool using nested supports,

$$\mathcal{S}_5 \subset \mathcal{S}_{10} \subset \mathcal{S}_{20}.$$

The query set remains fixed across shot settings and random seeds and is used only for final evaluation.

Four-class BNCI2014-001 uses Session 1 for calibration and Session 2 for evaluation. BNCI-LR follows the same cross-session split. For Cho2017, the first two runs form the calibration pool and all remaining runs form the query set. For Lee2019-MI, the offline training trials from Session 1 and Session 2 form the target calibration pool and query set, respectively.

Source pools. The final binary experimental pools contain 2,346 trials for BNCI-LR, 10,520 trials for Cho2017, and 10,800 trials for Lee2019-MI. For binary LODO evaluation, the two non-target datasets are combined to form the source pool. For four-class LOSO evaluation, the eight non-target BNCI2014-001 subjects form the source pool. Table 5 summarizes source-pool size and fixed-budget usage.

Table 5: Source-data usage relative to the source pool. Binary LODO uses the two non-target datasets, whereas four-class LOSO uses the eight non-target subjects.

Target task	Budget B	Eligible source pool	Trials selected	Usage
BNCI2014-001 (4-class)	500	4,608	500	10.85%
BNCI2014-001 (4-class)	1,000	4,608	1,000	21.70%
BNCI-LR	500	21,320	500	2.35%
Cho2017	500	13,146	500	3.80%
Lee2019-MI	500	12,866	500	3.89%

B Evaluation Metrics and Statistical Analysis

Let C denote the number of MI classes. All classification metrics are computed on the fixed query set of each held-out target subject. We use macro Balanced Accuracy (macro BA) as the primary metric because it assigns equal importance to every class and provides a consistent class-balanced measure across both the binary and four-class MI tasks. We additionally report per-class recall to reveal class-specific behavior and macro F1 to account for both false positives and false negatives.

Per-class Recall. For class c,

$$R_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}, \tag{13}$$

where TP_c and FN_c denote the numbers of correctly and incorrectly classified query trials belonging to class c, respectively. Reporting R_c separately helps determine whether performance gains are consistent across MI classes or driven by only a subset of them.

Macro Balanced Accuracy. Macro BA is defined as the mean class-wise recall:

$$\text{Macro BA} = \frac{1}{C}\sum_{c=1}^{C} R_c. \tag{14}$$

By weighting all classes equally, Macro BA is insensitive to differences in class frequency and avoids allowing performance on a dominant class to obscure poor performance on others. This property is particularly useful for comparing methods consistently across the binary and four-class evaluation settings considered in this work.

Macro F1. For each class c, precision and F1-score are computed in a one-versus-rest manner:

$$P_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c}, \qquad F_{1,c} = \frac{2P_c R_c}{P_c + R_c}, \tag{15}$$

where FP_c denotes trials from other classes incorrectly predicted as class c. Macro F1 is then

$$\text{Macro F1} = \frac{1}{C}\sum_{c=1}^{C} F_{1,c}. \tag{16}$$

Unlike Macro BA, which depends only on class-wise recall, macro F1 also penalizes false-positive predictions. It therefore provides a complementary class-balanced assessment of prediction quality by jointly considering precision and recall.

Aggregation and statistical analysis. Each shot-dependent configuration is evaluated using five shared random seeds. For each method and target subject, performance is first averaged across seeds and then summarized across subjects as mean $\pm$ sample standard deviation. The target subject, rather than the random seed, is treated as the paired statistical unit. For each binary LODO task, a Friedman omnibus test is conducted separately within each shot setting. TETHER is then compared with each competing method using two-sided paired Wilcoxon signed-rank tests across the target subjects of the corresponding dataset. Holm correction is applied over the nine pairwise comparisons separately within each dataset-shot condition. The shot-independent All-source reference is excluded from these tests.

C Baselines

We organize the baselines into four groups, each probing a different aspect of source-data selection (Killamsetty et al., 2021b;a). Random and unrestricted references compare target-conditioned source-trial retrieval with random sampling and exhaustive source reuse. Subject-level methods examine whether coarse source screening is sufficient. Trial-level similarity and graph methods compare pointwise target–source matching with pairwise relational modeling. Finally, transfer-aware methods provide EEG-specific comparisons with existing strategies designed to identify transferable source data.

C.1 Random and Unrestricted References

These references establish performance under selection without target information and under unrestricted source reuse, allowing us to separate the effect of learned source selection from that of simply using more data.

Class-balanced Random-B. This baseline uniformly samples B/C trials from each class of the complete source pool. It provides a budget- and class-matched reference without target conditioning and therefore measures whether a retrieval method exploits informative source–target relevance beyond random sampling.

All-source reference for four-class LOSO. For four-class LOSO evaluation, All-source reference uses all trials in the source pool together with the target support set and follows the same downstream decoding protocol as the fixed-budget methods. It represents exhaustive source reuse without source selection, providing a high-resource reference for assessing whether a compact selected subset can approach or exceed full-source training. Because its source-set size is not constrained by B, it is excluded from fixed-budget rankings.

All-source reference for binary LODO. For binary LODO evaluation, All-source uses all source trials together with the fixed reference support set $\mathcal{S}_T^{\mathrm{ref}}$. This shot-independent unrestricted reference evaluates transfer from the complete external source repository without target-conditioned source selection.

C.2 Subject-level Source Selection

These baselines evaluate whether coarse-grained source-domain relevance is sufficient for target adaptation. They exploit subject-level similarity while ignoring variation in transfer utility among individual trials from the same subject.

Subject-level Top-k + Random-B. This baseline ranks source subjects according to their class-conditional similarity to the target calibration set and retains the k highest-ranked subjects. A class-balanced subset of B trials is then sampled from the selected subjects. This design isolates the benefit of subject-level screening under the same trial budget as the fixed-budget methods. We use $k = 1$ for four-class $B = 500$, $k = 2$ for four-class $B = 1000$, and $k = 3$ for binary $B = 500$.

Subject-level Top-k + All Trials. This variant uses all available trials from the same k selected source subjects. It removes the fixed trial budget and therefore measures how far subject-level screening can benefit from retaining all data from the selected domains. Because its source-set size is not constrained by B, it is excluded from fixed-budget rankings.

C.3 Trial-level Similarity and Graph Baselines

These methods perform target-conditioned source-trial selection at the individual-trial level. They provide progressively richer comparisons, ranging from pointwise spectral and spatial similarity to pairwise relational propagation, and therefore test whether higher-order source organization provides benefits beyond conventional trial-level matching.

Bandpower Top-B. Bandpower Top-B represents each trial using its bandpower features and constructs a class-specific prototype from the target support trials. Source trials are ranked within each class by their distance to the corresponding target prototype, and the nearest B/C trials are retained. This baseline provides a simple and computationally inexpensive measure of target relevance based on conventional spectral information.

Riemannian Top-B. Riemannian Top-B represents each trial by its spatial covariance matrix. For each class, a Riemannian mean is computed from the target support trials. Source trials are ranked by their Riemannian distance to the corresponding target class mean, and the nearest B/C trials are retained. Compared with Bandpower Top-B, this baseline exploits geometry-aware spatial covariance structure while still evaluating source trials independently.

Pairwise kNN Graph Diffusion. This baseline constructs a pairwise kNN graph over source trials and target support trials. Labeled target nodes initialize class-conditioned relevance signals that are propagated through the graph to score source nodes, after which the B/C highest-scoring source trials from each class are retained. It provides a relational baseline beyond independent trial scoring and is particularly relevant for testing whether higher-order hypergraph structure offers benefits beyond pairwise graph propagation (Wang et al., 2023; Li et al., 2026a).

C.4 Transfer-aware Source Selection

These methods are representative EEG source-selection approaches that explicitly estimate source transferability or suppress potentially harmful source data. They provide domain-specific comparisons beyond generic similarity-based retrieval.

MTS-Top-B. MTS (Liang et al., 2021) estimates source-trial transferability on the symmetric positive-definite manifold. Its advantage is that transferability is assessed directly at the trial level in a geometry tailored to covariance-based EEG representations. Because the original method does not

necessarily return exactly B trials, MTS-Top-B ranks source trials by their MTS scores and retains the highest-scoring B/C trials from each class.

MEIS-Top-B. Our implementation of MEIS (Liang et al., 2024) first ranks source-subject domains by the Riemannian distance between their covariance centroids and that of the target support set. Domains are added in ascending distance order until the candidate pool contains at least B/C trials from every class. Within this pool, source trials are ranked by the closeness of their estimated kernel mean matching (KMM) weights to one, and the top B/C trials are retained per class.

RSSDA-Top-B. RSSDA (Ding et al., 2026) ranks source subjects using Riemannian within-class compactness and between-class separability. It provides a transfer-aware subject-level comparison in which source quality is evaluated using both class cohesion and discrimination rather than target similarity alone. Source subjects are considered in descending score order until the candidate pool contains enough trials, after which B/C trials are sampled from each class.

D Implementation Details

Signal preprocessing and feature extraction. All datasets follow a common signal-processing pipeline. Continuous EEG is first band-pass filtered at its native sampling rate using a fourth-order zero-phase Butterworth filter from 4 to $38\,\mathrm{Hz}$. For cross-dataset evaluation, recordings are then restricted to the 21 EEG channels shared exactly by BNCI2014-001, Cho2017, and Lee2019-MI: Fz, FC3, FC1, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, and POz. Channels follow the BNCI2014-001 ordering, and no channel interpolation or missing-electrode estimation is used. Common-average referencing is subsequently performed over the retained channels.

Signals are resampled to $250\,\mathrm{Hz}$ using polyphase resampling and cropped to 0.5–3.0 s after the MI cue, yielding 625 samples per channel. Each channel is independently z-score normalized within each trial using a denominator of $\mathrm{std} + 10^{-6}$. Before feature extraction, the resulting epochs are additionally zero-phase band-pass filtered at 8–$30\,\mathrm{Hz}$. Spatial covariance matrices are estimated using Oracle Approximating Shrinkage (OAS) (Chen et al., 2010), and spectral features are channel-wise log-bandpower in the 8–$13\,\mathrm{Hz}$ μ band and 13–$30\,\mathrm{Hz}$ β band. The same feature definitions are used by the retriever, the selection baselines, and the downstream decoder.

Shot sampling. All formal experiments use five random seeds, $\{2026, 2027, 2028, 2029, 2030\}$, and follow the nested support-set protocol described in Appendix A. The four-class experiments use $B \in \{500, 1000\}$, corresponding to $K_c \in \{125, 250\}$, whereas the binary experiments use $B = 500$ with $K_c = 250$.

Encoder and retriever training. The source encoder is first trained with source-label supervision and is fully frozen before retriever optimization. The encoder and retriever use independent Adam optimizers with the settings summarized in Table 7.

Retriever learning uses source-only pseudo-target episodes. Eligible pseudo-target domains contain at least 21 trials per class, so that the largest pseudo-support size leaves at least one pseudo-query trial. Eligible domains are deterministically ordered and visited cyclically across episodes, while the pseudo-support size follows

$$5 \to 10 \to 20 \to 5 \to \cdots .$$

Within each episode, trials from the held-out domain are shuffled using the corresponding seeded random-number generator. The first k_ε trials per class form the pseudo-support set and all remaining trials form the pseudo-query set. The episode-held-out source domain is included in the preceding source-encoder pretraining but excluded from the episode-specific dual-relation source hypergraph and source candidate pool. Each pseudo-target episode produces one retriever update, and a single retriever is trained across all three pseudo-support sizes and reused across target-shot settings.

Hyperparameter policy. Except for the downstream decoder parameters C_{LR} and ρ, the structural and retrieval hyperparameters are fixed a priori before formal evaluation rather than selected through target-side validation. They are subsequently kept fixed across target subjects, random seeds, and

Table 6: Key structural and utility-learning hyperparameters of TETHER.

Parameter	Value
Multi-view mixing weight η	0.70
Candidate neighbor pool k_{pool}	256
Relation neighbors ($k_{\text{loc}}/k_{\text{con}}$)	20/20
Cross-domain consensus coverage/cap ($U_{\min}/b$)	3/4
Initial retrieval depth M	64 ($K_c = 125$), 128 ($K_c = 250$)
Support attachments/temperature (m_{att}/τ_b)	8/0.20
HGNN layers/dropout	2/0.10
Utility parameters ($\tau_s/\gamma/\lambda_{\text{ridge}}$)	0.25/2.0/0.01

evaluation datasets. The only budget-dependent structural setting is the initial hyperedge-retrieval depth M, which is scaled with the class-wise source budget K_c: $M = 64$ for $K_c = 125$ and $M = 128$ for $K_c = 250$. This approximately preserves the initial candidate-retrieval density as the class-wise source budget increases.

The cross-domain consensus relation additionally enforces cross-domain coverage while limiting domination by a single source domain: each cross-domain consensus hyperedge must cover at least $U_{\min}$ source domains and contain at most b neighbors from any non-anchor domain. The remaining key structural and utility-learning settings are summarized in Table 6.

Decoder development and fitting. We instantiate the downstream decoder as an L2-regularized logistic regression classifier. Its parameters are selected once using calibration-only development experiments on four-class BNCI2014-001 with $B = 500$. For each target support set, 20% of the trials per class are reserved for validation, yielding $4/1$, $8/2$, and $16/4$ training/validation trials per class for the 5-, 10-, and 20-shot settings, respectively. We evaluate

$$C_{\text{LR}} \in \{0.01, 0.1, 1.0\}, \qquad \rho \in \{0.30, 0.50\},$$

where C_{LR} is its inverse regularization strength and ρ is the fraction of total training-weight mass assigned to target support trials.

Development uses seeds 2026 and 2027, all nine BNCI2014-001 target subjects, and all three shot settings, giving $2 \times 9 \times 3 = 54$ validation conditions. The configuration minimizing mean validation macro cross-entropy is

$$C_{\text{LR}} = 0.01, \qquad \rho = 0.50.$$

These values are frozen for all formal experiments, including both four-class source budgets, all three binary tasks, and the ablation experiments. Target query data are never used during hyperparameter development.

After source selection, the Riemannian reference and feature standardizers are refitted using only the selected source subset and then frozen. The resulting transformation is applied unchanged to the selected source trials, the complete target support set, and the target query set. The downstream decoder is fitted jointly on the selected source trials and the complete target support set using the frozen C_{LR} and ρ. The target query set is evaluated only after all preprocessing statistics, selections, and classifier parameters have been fixed.

E Complete Binary Cross-Dataset Results

E.1 Classification Performance

Tables 8–10 report the complete numerical results for the three binary LODO tasks with $B = 500$. We report per-class recall, Macro BA, and macro F1 for the 5-, 10-, and 20-shot settings. Bold and underlined values denote the best and second-best fixed-budget results, respectively; unrestricted references are excluded from ranking.

Table 7: Optimization settings for the source encoder, retriever, and downstream decoder.

Setting	Value
Encoder dimensions ($d_{\text{hid}}/d_{\text{emb}}$)	$128/64$
Encoder epochs/batch size	$20/256$
Encoder optimizer	Adam
Retriever updates	64
Encoder/retriever learning rate	10^{-3}
Encoder/retriever weight decay	10^{-4}
Retriever gradient clipping	5.0
Decoder solver/max iterations	L-BFGS/1000

Table 8: Binary MI performance on BNCI-LR.

Dataset	Shot	Method	Per-class Recall ↑		Macro BA ↑	Macro F1 ↑
			Left Hand	Right Hand		
BNCI-LR	5-shot	Class-balanced Random-B	**0.6850**±0.1041	0.5866±0.1948	0.6358±0.1264	0.6205±0.1391
		Bandpower Top-B	0.6116±0.1887	0.5509±0.2244	0.5813±0.1345	0.5433±0.1445
		Riemannian Top-B	0.5655±0.2215	0.7126±0.1597	0.6391±0.1703	0.5866±0.2035
		Subject-level Top-3 + All Trials	0.5592±0.2097	0.6462±0.1928	0.6027±0.1636	0.5781±0.1710
		Subject-level Top-3 + Random-B	0.4577±0.1746	0.6832±0.1944	0.5704±0.1189	0.5322±0.1174
		Pairwise kNN Graph Diffusion	0.6574±0.2122	0.6267±0.2347	0.6421±0.1690	0.6092±0.1905
		MTS-Top-B (Liang et al., 2021)	0.5426±0.1872	0.7185±0.1655	0.6306±0.1396	0.5937±0.1572
		MEIS-Top-B (Liang et al., 2024)	0.3232±0.3092	**0.9119**±0.0468	0.6175±0.1533	0.5301±0.2029
		RSSDA-Top-B (Ding et al., 2026)	0.6100±0.0790	0.5643±0.1177	0.5872±0.0911	0.5802±0.0945
		TETHER (Ours)	0.6722±0.1959	0.7414±0.1885	**0.7068**±0.1676	**0.6853**±0.1876
	10-shot	Class-balanced Random-B	0.6512±0.1391	0.6333±0.2105	0.6422±0.1347	0.6310±0.1415
		Bandpower Top-B	0.6391±0.1770	0.5941±0.2253	0.6166±0.1674	0.6023±0.1728
		Riemannian Top-B	0.5213±0.2038	0.7715±0.1531	0.6464±0.1691	0.6121±0.1795
		Subject-level Top-3 + All Trials	0.5985±0.2187	0.6626±0.1760	0.6306±0.1716	0.6054±0.1828
		Subject-level Top-3 + Random-B	0.5686±0.2093	0.6315±0.2013	0.6000±0.1589	0.5777±0.1678
		Pairwise kNN Graph Diffusion	0.6185±0.2363	0.6703±0.1992	0.6444±0.1768	0.6047±0.2068
		MTS-Top-B (Liang et al., 2021)	0.6277±0.1848	0.6810±0.1697	0.6543±0.1612	0.6376±0.1703
		MEIS-Top-B (Liang et al., 2024)	0.4349±0.2572	**0.8364**±0.0595	0.6356±0.1533	0.5778±0.1855
		RSSDA-Top-B (Ding et al., 2026)	0.6475±0.1201	0.5505±0.1415	0.5990±0.1148	0.5926±0.1165
		TETHER (Ours)	**0.6929**±0.1722	0.7688±0.1901	**0.7309**±0.1662	**0.7222**±0.1734
	20-shot	Class-balanced Random-B	0.6246±0.1629	0.6901±0.1778	0.6574±0.1603	0.6465±0.1674
		Bandpower Top-B	0.6244±0.1684	0.6101±0.2032	0.6173±0.1656	0.6086±0.1695
		Riemannian Top-B	0.6248±0.2472	0.7505±0.1760	0.6877±0.1992	0.6676±0.2127
		Subject-level Top-3 + All Trials	0.6310±0.2133	0.6895±0.1962	0.6603±0.1823	0.6415±0.1940
		Subject-level Top-3 + Random-B	0.6040±0.1681	0.6855±0.1867	0.6448±0.1688	0.6308±0.1757
		Pairwise kNN Graph Diffusion	0.6088±0.1750	0.6860±0.1884	0.6474±0.1719	0.6354±0.1786
		MTS-Top-B (Liang et al., 2021)	0.5687±0.2219	0.7344±0.1942	0.6516±0.1750	0.6292±0.1858
		MEIS-Top-B (Liang et al., 2024)	0.4317±0.3183	**0.8832**±0.0757	0.6575±0.1668	0.5934±0.2144
		RSSDA-Top-B (Ding et al., 2026)	0.6458±0.1490	0.6028±0.1417	0.6243±0.1326	0.6127±0.1398
		TETHER (Ours)	**0.7235**±0.1464	0.7731±0.1947	**0.7483**±0.1608	**0.7399**±0.1684
	–	All-source	0.7099±0.2509	0.6951±0.2376	0.7025±0.1714	0.6814±0.1889

E.2 Computational Efficiency

Table 11 reports the runtime of source selection and target adaptation on the three binary LODO tasks. For TETHER, Online only includes target-conditioned source-trial retrieval and final decoder fitting, whereas Full pipeline additionally includes the amortized one-time source-side retriever construction and training cost. The measurements exclude query-set forward inference.

Table 9: Binary MI performance on Cho2017.

Dataset	Shot	Method	Per-class Recall ↑		Macro BA ↑	Macro F1 ↑
			Left Hand	Right Hand		
Cho2017	5-shot	Class-balanced Random-B	0.5191±0.1106	0.6151±0.1230	0.5671±0.0860	0.5408±0.0892
		Subject-level Top-3 + Random-B	0.4923±0.1301	0.6306±0.1315	0.5614±0.0869	0.5376±0.0897
		Subject-level Top-3 + All Trials	0.4954±0.1336	0.6434±0.1296	0.5694±0.1010	0.5506±0.1054
		Bandpower Top-B	0.4595±0.1715	0.6491±0.1552	0.5543±0.0918	0.5045±0.1076
		Riemannian Top-B	0.4617±0.1498	0.6196±0.1482	0.5407±0.0870	0.4979±0.0945
		Pairwise kNN Graph Diffusion	0.5054±0.1657	0.5913±0.1868	0.5483±0.0831	0.4929±0.0973
		MTS-Top-B (Liang et al., 2021)	0.5296±0.1853	**0.6644**±0.1377	0.5970±0.1073	0.5695±0.1183
		MEIS-Top-B (Liang et al., 2024)	0.4842±0.1300	0.6142±0.1406	0.5492±0.0884	0.5246±0.0884
		RSSDA-Top-B (Ding et al., 2026)	0.5320±0.0955	0.5644±0.1018	0.5482±0.0694	0.5400±0.0703
		TETHER (Ours)	**0.6033**±0.1504	0.6313±0.1559	**0.6173**±0.1121	**0.5986**±0.1211
	10-shot	Class-balanced Random-B	0.4429±0.1181	**0.6832**±0.0930	0.5631±0.0839	0.5334±0.0901
		Subject-level Top-3 + Random-B	0.5218±0.1236	0.6123±0.1426	0.5670±0.1026	0.5518±0.1044
		Subject-level Top-3 + All Trials	0.5190±0.1504	0.6448±0.1447	0.5819±0.1128	0.5646±0.1192
		Bandpower Top-B	0.5153±0.1659	0.6059±0.1586	0.5606±0.0868	0.5340±0.0948
		Riemannian Top-B	0.4833±0.1592	0.6094±0.1549	0.5464±0.0969	0.5225±0.1010
		Pairwise kNN Graph Diffusion	0.4990±0.1627	0.6162±0.1457	0.5576±0.0995	0.5295±0.1061
		MTS-Top-B (Liang et al., 2021)	0.5619±0.1883	0.6206±0.1589	0.5912±0.1089	0.5657±0.1179
		MEIS-Top-B (Liang et al., 2024)	0.5178±0.1358	0.5952±0.1369	0.5565±0.0978	0.5399±0.0984
		RSSDA-Top-B (Ding et al., 2026)	0.5378±0.0943	0.5665±0.1007	0.5522±0.0703	0.5405±0.0734
		TETHER (Ours)	**0.6249**±0.1646	0.6463±0.1621	**0.6356**±0.1218	**0.6213**±0.1303
	20-shot	Class-balanced Random-B	0.5420±0.1394	0.6309±0.1325	0.5865±0.1168	0.5703±0.1223
		Subject-level Top-3 + Random-B	0.5441±0.1349	0.6139±0.1372	0.5790±0.1090	0.5653±0.1116
		Subject-level Top-3 + All Trials	0.5447±0.1534	0.6234±0.1375	0.5840±0.1237	0.5708±0.1279
		Bandpower Top-B	0.5343±0.1646	0.6484±0.1377	0.5913±0.1186	0.5699±0.1262
		Riemannian Top-B	0.4863±0.1307	0.6213±0.1577	0.5538±0.0979	0.5336±0.0964
		Pairwise kNN Graph Diffusion	0.5403±0.1661	0.6057±0.1497	0.5730±0.1193	0.5523±0.1263
		MTS-Top-B (Liang et al., 2021)	0.5801±0.1880	0.6276±0.1464	0.6039±0.1202	0.5808±0.1311
		MEIS-Top-B (Liang et al., 2024)	0.5306±0.1421	0.6090±0.1528	0.5698±0.1092	0.5532±0.1110
		RSSDA-Top-B (Ding et al., 2026)	0.5803±0.1094	0.5509±0.1208	0.5656±0.0845	0.5520±0.0894
		TETHER (Ours)	**0.6341**±0.1716	**0.6699**±0.1564	**0.6520**±0.1279	**0.6420**±0.1337
	–	All-source	0.6871±0.1929	0.6431±0.2096	0.6651±0.1238	0.6451±0.1391

Table 10: Binary motor-imagery performance on Lee2019-MI.

Dataset	Shot	Method	Per-class Recall ↑		Macro BA ↑	Macro F1 ↑
			Left Hand	Right Hand		
Lee2019-MI	5-shot	Class-balanced Random-B	0.5012±0.2106	0.6475±0.1755	0.5744±0.1066	0.5596±0.1123
		Subject-level Top-3 + Random-B	0.6188±0.2366	0.5144±0.2218	0.5666±0.1154	0.5452±0.1256
		Subject-level Top-3 + All Trials	0.6600±0.2331	0.4875±0.2543	0.5738±0.1201	0.5456±0.1381
		Bandpower Top-B	0.6000±0.2966	0.5088±0.3059	0.5544±0.0965	0.5062±0.1306
		Riemannian Top-B	**0.7425**±0.2022	0.4838±0.3501	0.6131±0.1345	0.5664±0.1777
		Pairwise kNN Graph Diffusion	0.7144±0.2932	0.4631±0.3590	0.5888±0.1306	0.5287±0.1717
		MTS-Top-B (Liang et al., 2021)	0.5065±0.2080	0.6733±0.1682	0.5899±0.1085	0.5571±0.1209
		MEIS-Top-B (Liang et al., 2024)	0.6421±0.1824	0.5201±0.1956	0.5811±0.1160	0.5484±0.1253
		RSSDA-Top-B (Ding et al., 2026)	0.5799±0.1185	0.5550±0.1319	0.5674±0.0960	0.5569±0.0984
		TETHER (Ours)	0.6128±0.1975	**0.6856**±0.1934	**0.6492**±0.1287	**0.6196**±0.1454
	10-shot	Class-balanced Random-B	0.4806±0.2271	0.6462±0.1854	0.5634±0.1002	0.5436±0.1100
		Subject-level Top-3 + Random-B	0.5562±0.2389	0.5475±0.2429	0.5519±0.1026	0.5276±0.1128
		Subject-level Top-3 + All Trials	0.5869±0.2620	0.5744±0.2661	0.5806±0.1269	0.5535±0.1414
		Bandpower Top-B	0.5550±0.2588	0.6000±0.2087	0.5775±0.1316	0.5561±0.1487
		Riemannian Top-B	0.6231±0.2369	0.5731±0.2498	0.5981±0.1342	0.5788±0.1417
		Pairwise kNN Graph Diffusion	0.6006±0.2897	0.5831±0.3098	0.5919±0.1164	0.5501±0.1484
		MTS-Top-B (Liang et al., 2021)	0.5534±0.1835	0.6431±0.1453	0.5983±0.1154	0.5783±0.1223
		MEIS-Top-B (Liang et al., 2024)	**0.6672**±0.1733	0.5094±0.1999	0.5883±0.1261	0.5599±0.1375
		RSSDA-Top-B (Ding et al., 2026)	0.5801±0.1264	0.5547±0.1279	0.5674±0.0952	0.5538±0.0981
		TETHER (Ours)	0.6244±0.2048	**0.6902**±0.2116	**0.6573**±0.1357	**0.6294**±0.1527
	20-shot	Class-balanced Random-B	0.5487±0.2098	0.6531±0.1467	0.6009±0.1299	0.5925±0.1346
		Subject-level Top-3 + Random-B	0.5856±0.2129	0.5562±0.2214	0.5709±0.1238	0.5545±0.1328
		Subject-level Top-3 + All Trials	**0.6538**±0.2348	0.5412±0.2573	0.5975±0.1416	0.5745±0.1582
		Bandpower Top-B	0.5556±0.2736	0.6438±0.2290	0.5997±0.1388	0.5755±0.1565
		Riemannian Top-B	0.5650±0.2585	0.6450±0.2871	0.6050±0.1444	0.5769±0.1619
		Pairwise kNN Graph Diffusion	0.6413±0.2526	0.5877±0.2850	0.6145±0.1314	0.5866±0.1504
		MTS-Top-B (Liang et al., 2021)	0.5650±0.2163	0.6814±0.1585	0.6232±0.1346	0.6034±0.1443
		MEIS-Top-B (Liang et al., 2024)	0.6436±0.1816	0.5676±0.2016	0.6056±0.1365	0.5795±0.1462
		RSSDA-Top-B (Ding et al., 2026)	0.5812±0.1361	0.5685±0.1289	0.5749±0.0974	0.5606±0.1022
		TETHER (Ours)	0.6380±0.2143	**0.7006**±0.2263	**0.6693**±0.1388	**0.6408**±0.1575
	–	All-source	0.4906±0.2536	0.9062±0.0905	0.6984±0.1286	0.6706±0.1540

Table 11: Runtime comparison on the binary tasks. Avg. denotes seconds per subject–shot–seed condition, whereas Total denotes aggregate serial hours. Full pipeline includes the amortized offline retriever cost; Online only reports target-conditioned source-trial selection and adaptation.

	BNCI-LR		Cho2017		Lee2019-MI	
Method	**Avg. (s)**	**Total (h)**	**Avg. (s)**	**Total (h)**	**Avg. (s)**	**Total (h)**
Class-balanced Random-B	7.511	0.282	5.971	1.294	6.547	1.473
Bandpower Top-B	6.257	0.235	5.959	1.291	5.747	1.293
Riemannian Top-B	8.427	0.316	6.755	1.464	7.423	1.670
Subject-level Top-3 + Random-B	6.669	0.250	5.991	1.298	6.203	1.396
Subject-level Top-3 + All Trials	8.038	0.301	7.041	1.526	7.380	1.661
Pairwise kNN Graph Diffusion	10.205	0.383	8.057	1.746	8.307	1.869
MTS-Top-B	156.561	5.871	98.061	21.247	96.434	21.698
MEIS-Top-B	97.815	3.668	61.839	13.398	60.983	13.721
RSSDA-Top-B	50.197	1.882	43.172	9.354	40.567	9.128
All-source	20.886	0.261	33.352	2.409	29.803	2.235
TETHER (Full pipeline)	**5.107**	**0.192**	**1.295**	**0.281**	**1.243**	**0.280**
TETHER (Online only)	**0.646**	**0.024**	**0.647**	**0.140**	**0.641**	**0.144**