MGT: MULTI-GRANULARITY TRANSFER FOR CROSS-SUBJECT EEG ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Building EEG models that can adapt reliably to unseen subjects across diverse paradigms remains challenging because EEG signals vary substantially across subjects and tasks. Although existing approaches increasingly emphasize transferable EEG representations, knowledge acquired from different sources does not contribute equally when adapting to unseen subjects, and indiscriminate transfer inevitably induces negative transfer and degrades cross-subject generalization. We therefore propose Multi-Granularity Transfer (MGT), a general transfer framework for cross-subject EEG adaptation. MGT jointly models transfer relationships in heterogeneous EEG knowledge at multiple granularities, enabling the model to preserve universally transferable patterns while suppressing domain-mismatched knowledge. We further introduce a progressive transfer mechanism to stabilize knowledge adaptation as learning proceeds, thereby improving the robustness of cross-subject adaptation. We systematically evaluate MGT on seven EEG datasets spanning five paradigms, including motor imagery, emotion recognition, and imagined speech. Under cross-subject evaluation protocols, MGT consistently achieves robust decoding performance across these diverse paradigms. The results demonstrate the effectiveness of modeling transferability at multiple granularities and highlight the potential of MGT for cross-subject EEG decoding.

1 INTRODUCTION

Electroencephalography (EEG) decoding spans a range of paradigms, including motor imagery, emotion recognition, imagined speech, rapid serial visual presentation, and error-related negativity. Yet models trained on one group of subjects often generalize poorly to unseen subjects because EEG signals vary substantially across individuals. Cross-subject methods usually address this problem by learning domain-invariant representations or adapting a model to reduce discrepancies between source and target subjects. HEEGNet (Li et al., 2026) learns domain-invariant hyperbolic representations with a coarse-to-fine adaptation strategy, whereas BTTA-DG (Luo et al., 2026) performs test-time adaptation to handle cross-subject and cross-session shifts in motor-imagery EEG. These approaches improve cross-subject adaptation, but they do not explicitly model whether knowledge acquired from different sources contributes equally to an unseen subject. In practice, transferability may vary across source subjects and individual source samples, indicating heterogeneity at both the source and sample levels.

This heterogeneity makes uniform source utilization potentially problematic. Prior work on cross-subject EEG has shown that transfer can depend on the relationship between source and target subjects. Wang et al. (2024b) use dynamic transfer to match source and target samples with higher similarity in the latent space, while Cui et al. (2025) adapt knowledge sharing according to target-specific inter-subject differences. These studies suggest that knowledge acquired from different sources may not contribute equally when adapting to a particular target subject. The same issue can arise within a source, where individual samples and local structures may differ in their compatibility with the target distribution. Uniform source utilization may therefore retain domain-mismatched knowledge, while a single-granularity estimate can overlook local variation in transferability. We hypothesize that cross-subject EEG adaptation benefits from modeling transferability at both the source and sample levels rather than treating all source knowledge uniformly.

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.

To test this hypothesis, we propose Multi-Granularity Transfer (MGT) for cross-subject EEG adaptation. MGT models transferability at two complementary levels. At the source level, coarse-grained transfer estimates the overall compatibility between each source subject and the target. At the sample level, fine-grained transfer further identifies source samples whose discriminative or local structural characteristics are mismatched with the target distribution. The resulting transfer factors are then combined to regulate how different source knowledge contributes to model learning. We further introduce a progressive transfer mechanism that gradually incorporates these transfer factors during training, allowing optimization to move from uniform source utilization toward selective transfer as learning proceeds.

Our contributions are threefold:

- We formulate source transferability at two complementary granularities and propose MGT, a general transfer framework for cross-subject EEG adaptation that characterizes both source-level compatibility and sample-level variation in transferability.
- A progressive transfer mechanism is developed to gradually incorporate the estimated transfer factors during optimization, reducing abrupt changes in source contribution during early training.
- MGT is evaluated under cross-subject protocols on seven EEG datasets spanning five paradigms to examine whether the same transfer mechanism can operate across heterogeneous EEG settings.

2 METHODOLOGY

2.1 MGT OVERVIEW

Let $\mathcal{D}_{\mathrm{src}} = \{(x_i^s, y_i^s, s(i))\}_{i=1}^{N_s}$ denote the pooled labeled source data, where $s(i) \in \{1, \dots, S\}$ identifies the source subject, and let $\mathcal{D}_t = \{(x_j^t, y_j^t)\}_{j=1}^{n_t}$ denote the labeled target calibration set. Source and target data within each benchmark share the label space $\mathcal{Y} = \{1, \dots, K\}$.

Let F_θ denote the EEG representation model, which maps an input x to $H = F_\theta(x) \in \mathbb{R}^{L \times d}$, where L is the number of latent tokens and d their dimension. We instantiate F_θ with GAT (Song et al., 2023).

An overview of MGT is shown in Fig. 1. MGT first estimates source- and sample-level transferability in a frozen reference space, fuses the resulting transfer factors, and progressively incorporates them into final-model optimization.

Reference-space pretraining. Transfer factors are estimated before final-model training. We first train a reference model on $\mathcal{D}_{\mathrm{src}}$ and $\mathcal{D}_t$ with uniform source contributions, and then freeze its encoder. For an input x, we obtain the normalized reference representation

$$z(x) = \frac{1}{L} \sum_{\ell=1}^{L} H_{\mathrm{ref}}(x)_{\ell,:}, \qquad \bar{z}(x) = \frac{z(x)}{\|z(x)\|_2 + \epsilon}, \tag{1}$$

where $\epsilon > 0$ is used for numerical stability. The separate reference encoder keeps transferability estimation fixed rather than letting the estimated factors change with the final model they regulate. The final model is initialized independently.

2.2 MULTI-GRANULARITY TRANSFER

2.2.1 COARSE-GRAINED TRANSFER

For source subject s and class c, define $Z_{s,c} = \{\bar{z}_i^s \mid y_i^s = c,\ s(i) = s\}$ and $Z_{t,c} = \{\bar{z}_j^t \mid y_j^t = c\}$. We measure their discrepancy using entropy-regularized optimal transport (Courty et al., 2017). For two representation sets $A = \{a_u\}$ and $B = \{b_v\}$, the discrepancy is $D_\eta(A, B) = \langle \Pi_\eta^*, M \rangle$, where $M_{uv} = \|a_u - b_v\|_2^2$ and Π_η^* is obtained by Sinkhorn iterations (Cuturi, 2013). The source-level discrepancy is

$$\Delta_s = \frac{1}{|\mathcal{Y}_s|} \sum_{c \in \mathcal{Y}_s} D_\eta(Z_{s,c}, Z_{t,c}), \tag{2}$$

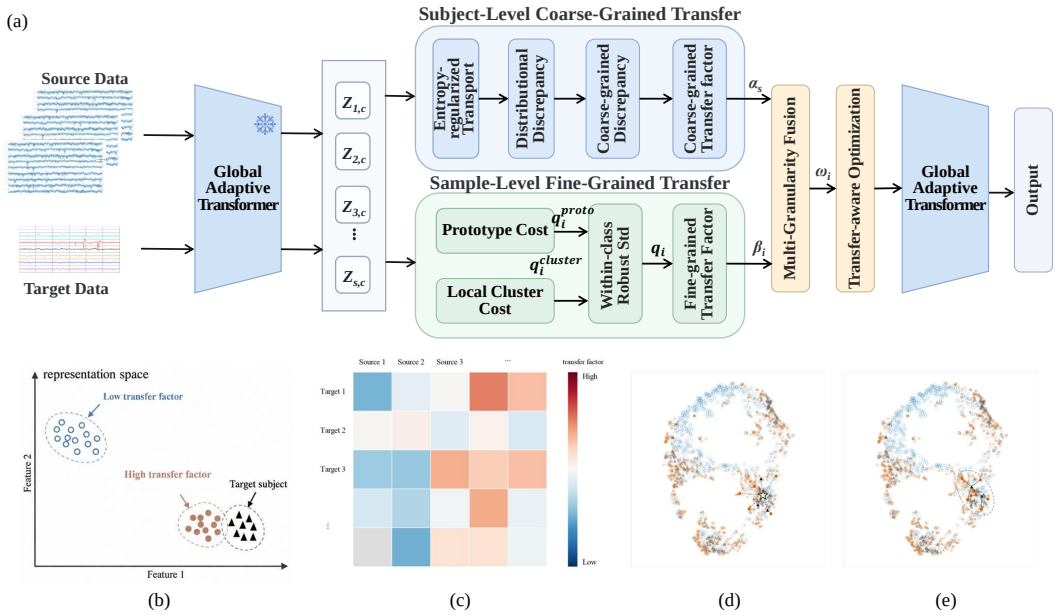


Figure 1: Overview of the proposed Multi-Granularity Transfer (MGT) framework. (a) MGT estimates coarse- and fine-grained transfer factors in a frozen reference space, fuses them, and progressively regulates source contributions during final-model optimization. (b–c) Coarse-grained transfer characterizes source–target compatibility at the subject level. (d–e) Fine-grained transfer measures sample-level compatibility through target-class prototype and local-structure costs.

where $\mathcal{Y}_s$ contains the classes shared by source s and the target. We convert Δ_s into the coarse-grained transfer factor

$$\alpha_s = \mathrm{clip}\left(\frac{\exp(-\Delta_s/\tau_s)}{S^{-1}\sum_{r=1}^{S}\exp(-\Delta_r/\tau_s)+\epsilon}, \alpha_{\min}, \alpha_{\max} \right), \tag{3}$$

where r indexes the source subjects, τ_s is the median of $\{\Delta_r\}_{r=1}^{S}$, and $\alpha_{\min}$ and $\alpha_{\max}$ bound the range of source-level modulation. A smaller Δ_s gives a larger α_s.

2.2.2 Fine-Grained Transfer

Samples from the same source share α_s, but their relationship with the target can still differ. We estimate this variation using a prototype cost and a local structural cost.

For class c, the target prototype is

$$p_c^t = \frac{1}{n_{t,c}} \sum_{j:y_j^t=c} \bar{z}_j^t, \tag{4}$$

where $n_{t,c}$ is the number of target calibration samples in class c. For source sample i, we define

$$q_i^{\mathrm{proto}} = \frac{\|\bar{z}_i^s - p_{y_i^s}^t\|_2^2}{\min_{c\neq y_i^s}\|\bar{z}_i^s - p_c^t\|_2^2 + \epsilon}. \tag{5}$$

A smaller value means that the sample is closer to its corresponding target class than to competing classes.

Prototype matching treats each source sample independently. We therefore also cluster the normalized source representations within each class. Let $I_c = \{i : y_i^s = c\}$ and set

$$K_c = \left\lceil \frac{|I_c|}{m} \right\rceil,$$

where m controls the nominal cluster size and is fixed to 64 across all experiments to provide a moderate clustering granularity. We use K-means, and MiniBatch K-means for classes containing at least 512 samples (Sculley, 2010). Let G_i denote the cluster containing sample i. Its local structural

cost is

$$q_i^{\text{cluster}} = \frac{D_\eta(G_i, Z_{t,y_i^s})}{\min_{c \neq y_i^s} D_\eta(G_i, Z_{t,c}) + \epsilon}. \tag{6}$$

The two costs are combined as

$$q_i = \lambda_{\text{proto}} q_i^{\text{proto}} + \lambda_{\text{cluster}} q_i^{\text{cluster}}, \tag{7}$$

where λ_{proto} and λ_{cluster} control their relative contributions.

Because the scale of q_i can differ across classes, we normalize it within each class using the median and median absolute deviation (MAD): $m_c = \text{median}_{i \in I_c}\, q_i$ and $r_c = \text{median}_{i \in I_c}\, |q_i - m_c|$. The normalized cost is $\tilde{q}_i = \frac{q_i - m_{y_i^s}}{r_{y_i^s} + \epsilon}$. The fine-grained transfer factor is then

$$\beta_i = \beta_{\min} + (1 - \beta_{\min}) \exp\left(-\frac{\max(0, \tilde{q}_i)}{\tau_t} \right), \tag{8}$$

where $\beta_{\min}$ sets the lower bound and τ_t controls the attenuation rate.

2.2.3 Multi-Granularity Fusion

We fuse the source- and sample-level factors multiplicatively:

$$u_i = \alpha_{s(i)} \beta_i. \tag{9}$$

With multiplication, a low compatibility estimate at either granularity lowers the contribution of sample i; a large fused factor requires both source-level and sample-level compatibility.

To avoid changing the empirical class balance, the fused factors are normalized within each class and then clipped:

$$w_i = \text{clip}\left(\frac{u_i}{|I_{y_i^s}|^{-1} \sum_{k \in I_{y_i^s}} u_k}, w_{\min}, w_{\max} \right), \tag{10}$$

where $w_{\min}$ and $w_{\max}$ bound the final transfer factor. Before clipping, each class has a mean factor of one.

2.3 Progressive Transfer Mechanism

The factors above are estimated from the frozen reference encoder, whereas the final representation model starts from a new initialization. Applying the full factors from the first epoch would therefore impose the reference-space transfer pattern before the final representation has developed. We instead introduce the factors progressively:

$$\rho_e = \min\left(1, \frac{e}{E_{\text{ramp}}}\right), \qquad w_i^{(e)} = (1 - \rho_e) + \rho_e w_i, \tag{11}$$

where e is the training epoch and E_{ramp} is the ramp duration. Thus $w_i^{(0)} = 1$, and the effective factor reaches w_i after E_{ramp} epochs. The estimated w_i itself remains fixed during final training.

Transfer-aware optimization. The effective factor $w_i^{(e)}$ is applied to source-dependent terms in the final objective. For a source mini-batch $\mathcal{B}_s$, we use

$$\mathcal{A}_s^{(e)}[v_i] = \frac{\sum_{i \in \mathcal{B}_s} w_i^{(e)} v_i}{\sum_{i \in \mathcal{B}_s} w_i^{(e)} + \epsilon}, \tag{12}$$

where v_i denotes a source-dependent loss term. In our GAT implementation, this aggregation replaces the ordinary source average in classification, source–target alignment, and class-structure regularization, while target terms remain unweighted. The inherited GAT objectives, target augmentation, and optimization details are provided in Appendix B.

At inference, only the final representation model and classifier are retained; source EEG data and target calibration labels are not required.

3 EXPERIMENTS

3.1 EXPERIMENTAL SETUP

Datasets. We evaluate MGT on seven public EEG datasets spanning five paradigms: motor imagery (MI), emotion recognition, imagined speech, rapid serial visual presentation (RSVP), and error-related negativity (ERN). Their main statistics and target calibration protocols are summarized in Table 1. Original dataset sources include BCI Competition IV (Tangermann et al., 2012), OpenBMI (Lee et al., 2019), SEED (Zheng & Lu, 2015), and the imagined-speech data (Nguyen et al., 2018). Recent ICLR and affiliated-workshop research spans continual EEG learning, cross-subject emotion recognition, EEG representation models, and foundation-model benchmarking (Zhou et al., 2025; Liu et al., 2024; Jiang et al., 2024; Wang et al., 2025; Yang et al., 2026; Nguyen et al., 2026).

Table 1: Summary of the seven EEG datasets used in our experiments. Target calibration denotes the labeled target data available during adaptation.

Dataset	Paradigm	Subjects	Classes	EEG ch.	Sampling rate (Hz)	Target calibration
BCI-IV-2a	Motor imagery	9	4	22	250	288 trials
BCI-IV-2b	Motor imagery	9	2	3	250	400 trials
OpenBMI	Motor imagery	54	2	62	$1{,}000 \rightarrow 250$	100 trials (Session 1)
SEED	Emotion recognition	15	3	62	$1{,}000 \rightarrow 200$	Sessions 1–2
NM000252	Imagined speech	6	2	60	$1{,}000 \rightarrow 256$	160 trials (480 windows)
RSVP	RSVP	11	2	8	2,048	294–452 trials
ERN	ERN	16	2	56	$600 \rightarrow 200$	240 trials

Because the datasets differ in session structure and available labeled target data, we use dataset-specific calibration sets rather than enforcing a fixed calibration size. The calibration protocol for each benchmark follows its corresponding labeled session, trial, or window partition, as summarized in Table 1.

Protocol and implementation. For each dataset, one subject is treated as the target and the remaining subjects form the source pool. The labeled target calibration data are used for adaptation and transferability estimation, while a separate target partition is used for evaluation. We use GAT as the representation model and pretrain the reference model for 80 epochs. Progressive transfer is introduced over the first 100 epochs of final training. Adam with a learning rate of 2×10^{-4} and a batch size of 64 is used for all datasets. Dataset-specific preprocessing and remaining implementation details are provided in the Appendix.

Evaluation metrics. For all seven datasets, we report subject-averaged balanced accuracy together with its subject-wise standard deviation. For BCI-IV-2a, BCI-IV-2b, OpenBMI, SEED, and NM000252, we additionally report Cohen's Kappa; for RSVP and ERN, we additionally report AUROC.

3.2 MAIN RESULTS

As shown in Table 2, MGT achieves the best primary-metric performance on all six datasets reported in the main text. The improvements are particularly clear on RSVP, BCI-IV-2a, and SEED. MGT also achieves the best Kappa on BCI-IV-2a, OpenBMI, SEED, and NM000252, and the best AUROC on RSVP. Results on BCI-IV-2b, the seventh dataset included in our evaluation, are reported in Appendix D.1. GAT (official updated) uses the authors' subsequently released v0.72 implementation; its differences from the originally reported GAT are detailed in Appendix C.2.

3.3 ABLATION STUDY

We examine the three main components of MGT on BCI-IV-2a, SEED, and NM000252. Starting from Full MGT, we remove Coarse-Grained Transfer, Fine-Grained Transfer, and Progressive Transfer independently while keeping the remaining training configuration unchanged. Without Progressive Transfer, the estimated transfer factors are applied directly throughout final training.

Table 2: Comparison of cross-subject EEG decoding performance on six representative datasets.

Methods	BCI Competition IV 2a (4-Class)			Methods	OpenBMI (2-Class)		
	BAcc	Kappa	Std		BAcc	Kappa	Std
MIN2Net	60.03	0.4671	9.24	Pooled CSP	65.65	0.3130	16.11
FBCSP	67.75	0.5700	13.73	ShallowConvNet	70.13	0.4026	12.71
EEGNet	70.22	0.6633	10.72	EEGNet	73.81	0.4763	14.73
ShallowConvNet	72.53	0.6338	13.18	Spectral-Spatial CNN	74.15	0.4830	15.83
DRDA	74.74	0.6633	12.96	GAT	74.63	0.4926	13.62
GAT	76.58	0.6877	15.98	MIN2Net	75.78	0.5156	14.52
BDAN-SPD	77.49	0.6998	15.22	BDAN-SPD	79.37	0.5874	12.99
GAT (official updated)	81.17	0.7490	11.27	GAT (official updated)	82.20	0.6441	13.06
Ours	**83.18**	**0.7757**	**10.16**	**Ours**	**82.65**	**0.6530**	**12.59**
Methods	**SEED (3-Class)**			**Methods**	**NM000252 (2-Class)**		
	BAcc	Kappa	Std		BAcc	Kappa	Std
CDCN	69.30	0.5400	9.40	RF	52.74	0.0548	5.10
MLP	70.10	0.5500	9.40	sLDA	55.55	0.1110	7.32
STFFNN	72.00	0.5790	9.00	Bi-RNN	55.90	0.1180	7.87
STRNN	73.20	0.5980	11.10	AdaBN	59.41	0.1882	5.84
3D-CNN	74.20	0.6130	7.80	CDAN	60.49	0.2098	6.92
MMResLSTM	74.40	0.6160	10.40	MMD	60.81	0.2162	5.80
MetaEmotionNet	77.50	0.6630	8.80	VADA	61.94	0.2388	5.67
GAT (official updated)	79.84	0.6982	13.87	GAT (official updated)	61.94	0.2389	6.14
Ours	**81.44**	**0.7218**	**13.10**	**Ours**	**62.78**	**0.2556**	**6.00**
Methods	**RSVP (2-Class)**			**Methods**	**ERN (2-Class)**		
	BAcc	AUROC	Std		BAcc	AUROC	Std
xDAWN-SVM	61.37	62.56	0.00	CBraMod	50.18	52.98	0.00
MEKT-R	68.38	--	6.36	BIOT	51.18	54.95	0.89
EA-xDAWN-SVM	69.07	--	9.05	LaBraM	54.39	56.93	0.29
DANN	72.51	73.17	7.89	STELAR-base	55.20	63.25	0.00
ATLDCN	72.71	74.07	5.44	BENDR	56.72	60.30	0.20
SFSA	72.94	74.22	8.27	EEGPT	58.37	66.21	0.64
CST-TVA-DRTL	73.95	74.48	5.96	BioCodec	60.90	64.90	0.70
GAT (official updated)	83.70	90.70	8.17	GAT (official updated)	78.89	**83.99**	10.80
Ours	**86.05**	**91.58**	**7.22**	**Ours**	**79.09**	83.97	**10.27**

Table 3: Ablation results on BCI-IV-2a, SEED, and NM000252.

Dataset	Variant	BAcc. (%) ↑	Kappa ↑
BCI-IV-2a	w/o Coarse-grained Transfer	81.79	0.7572
	w/o Fine-grained Transfer	82.14	0.7618
	w/o Progressive Transfer	81.79	0.7572
	Full MGT	**83.18**	**0.7757**
SEED	w/o Coarse-grained Transfer	77.29	0.6600
	w/o Fine-grained Transfer	77.63	0.6650
	w/o Progressive Transfer	76.98	0.6551
	Full MGT	**81.44**	**0.7218**
NM000252	w/o Coarse-grained Transfer	62.64	0.2528
	w/o Fine-grained Transfer	60.83	0.2167
	w/o Progressive Transfer	61.94	0.2389
	Full MGT	**62.78**	**0.2556**

Table 3 shows lower balanced accuracy and Kappa for every ablation than Full MGT on each dataset. The size of the decrease varies across paradigms: removing Progressive Transfer has the largest effect on SEED, whereas removing Fine-Grained Transfer has the largest effect on NM000252.

These results support the complementary roles of source- and sample-level transferability and the progressive transfer schedule.

3.4 TRANSFER ANALYSIS

We examine sample-level transferability, class-wise decoding, source-level compatibility, and parameter sensitivity on BCI-IV-2a.

Sample-level transfer patterns. We further examine how the fused transfer factors relate to the proximity of individual source samples to the target distribution. Source samples with transfer factors at or above the within-class 70th percentile and at or below the 30th percentile are treated as high- and low-factor samples, respectively. Figure 2 visualizes these samples together with the target calibration samples for representative target subjects S01, S04, and S07.

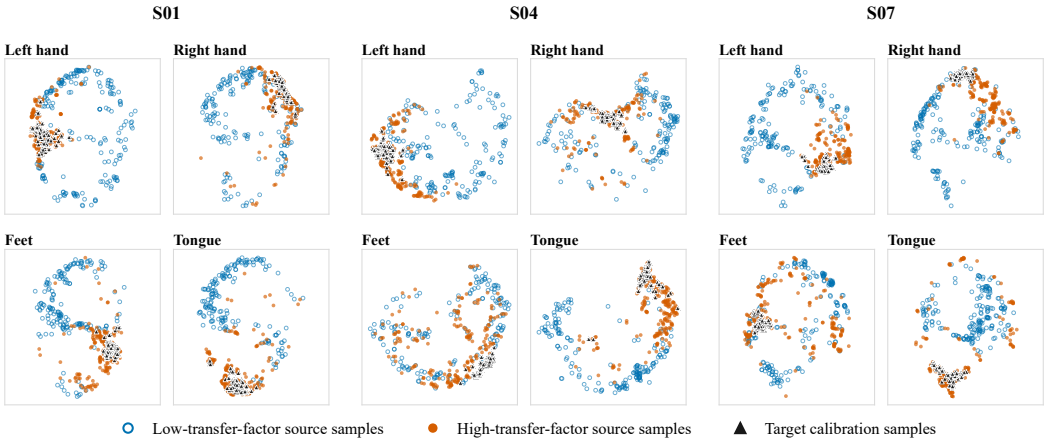


Figure 2: Sample-level transfer patterns on BCI Competition IV-2a. t-SNE (van der Maaten & Hinton, 2008) embeddings of mean-pooled 40-dimensional final-model representations are shown for representative target subjects S01, S04, and S07 across the four motor-imagery classes. Blue hollow and orange filled circles denote source samples with fused transfer factors at or below the within-class 30th percentile and at or above the 70th percentile, respectively; black triangles denote target calibration samples. Source samples with intermediate transfer factors are omitted from display. High-factor source samples show greater spatial overlap with the corresponding target calibration samples, whereas low-factor samples are more dispersed.

To quantify this pattern beyond the two-dimensional visualization, we use all nine target subjects and compute Euclidean distances in the original 40-dimensional final-model representation space. For each target subject and class, we compute the distance from each source sample to the corresponding target-class centroid and average the distances separately within the high- and low-factor groups.

Table 4: Euclidean distance from high- and low-transfer-factor source samples to the corresponding target-class centroid on BCI-IV-2a. Distances are computed in the 40-dimensional final-model representation space. Results are reported as mean ± standard deviation: class-wise statistics are computed over nine target subjects, and the overall statistics over all 36 target-subject/class pairs.

Class	High factor ↓	Low factor ↓
Left hand	0.686 ± 0.461	1.324 ± 0.305
Right hand	0.693 ± 0.175	1.277 ± 0.137
Feet	0.890 ± 0.291	1.443 ± 0.162
Tongue	0.705 ± 0.318	1.409 ± 0.220
Overall	$\mathbf{0.743 \pm 0.325}$	$\mathbf{1.363 \pm 0.217}$

High-factor samples have smaller target-class distances than low-factor samples for all four classes, with overall distances of 0.743 ± 0.325 and 1.363 ± 0.217, respectively. This quantitative result agrees with the spatial pattern observed in Fig. 2. Extended visualizations for all nine target subjects are provided in Appendix E.1.

Class-wise decoding patterns. Figure 3 shows recalls of 83.64%, 84.41%, 87.19%, and 77.47% for left-hand, right-hand, feet, and tongue imagery, respectively. Tongue imagery remains com-

paratively more challenging, with its largest residual confusions occurring with feet (9.10%) and right-hand imagery (8.33%). Subject-wise confusion matrices for all nine target subjects are provided in Appendix E.2.

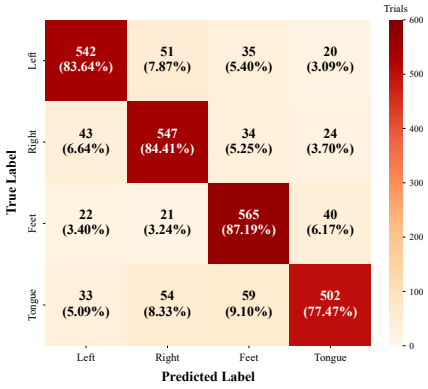
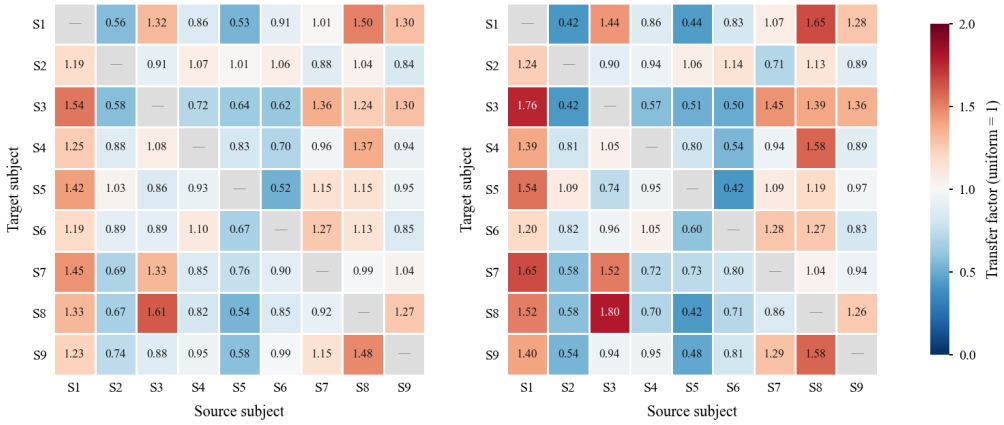


Figure 3: Class-wise confusion matrix of MGT on BCI Competition IV-2a over 2,592 target trials from nine subjects. Cells report trial counts and row-normalized percentages.

Source-level transfer patterns. Figure 4 shows the coarse-grained transfer factors $\alpha_{s\to t}$ and the mean fused transfer factors $\bar{w}_{s\to t}$ across source–target subject pairs. For a fixed target, different source subjects receive factors both above and below the uniform value of one, and the pattern changes across targets. Thus, source-level compatibility is target dependent rather than uniform. After sample-level modulation, the mean fused factors remain heterogeneous, further adjusting the contribution of individual sources.



(a) Coarse-grained transfer factor $\alpha_{s\to t}$ (b) Mean fused transfer factor $\bar{w}_{s\to t}$

Figure 4: Source–target transfer factors on BCI-IV-2a. A value of 1 denotes uniform source contribution; diagonal entries are omitted because the target subject is excluded from the source pool.

Multi-granularity transfer-factor distributions. Figure 5 further examines the source- and sample-level heterogeneity underlying MGT. For a fixed target subject, transfer-factor distributions vary substantially across source subjects, with different sources receiving predominantly enhanced or attenuated contributions. Meanwhile, individual samples within the same source can span a considerably wider range of transfer factors. This simultaneous across-source and within-source variation provides direct evidence for modeling transferability at both source and sample granularities.

Parameter sensitivity. On BCI-IV-2a, we vary λ_{proto} from 0.1 to 0.9 while setting $\lambda_{\text{cluster}} = 1 - \lambda_{\text{proto}}$. The absolute performance values in the sensitivity analysis differ from the main results because, for computational efficiency, these experiments used a reduced training protocol with early stopping enabled after 200 epochs. Balanced accuracy peaks at 80.98% when the prototype and cluster terms are equally weighted and decreases as the balance shifts toward either term (Fig. 6a).

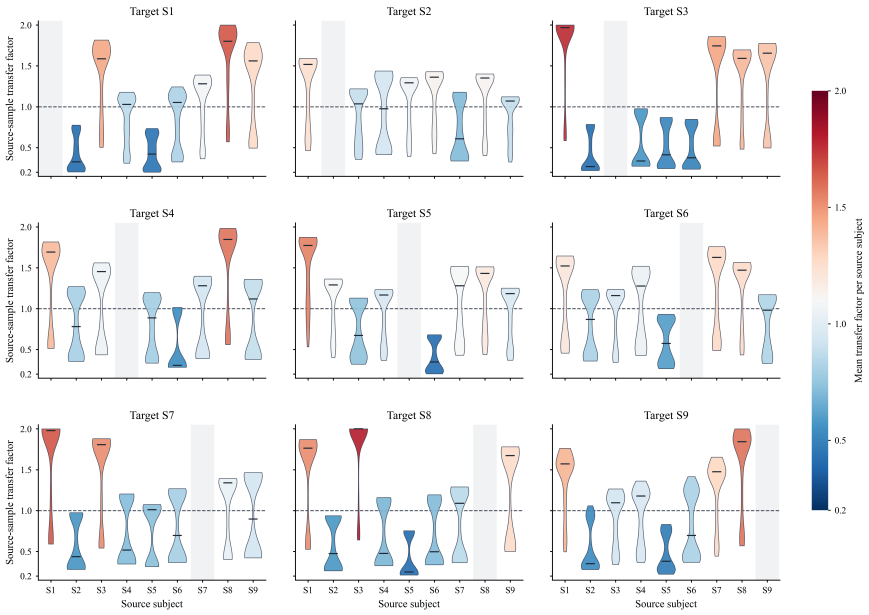
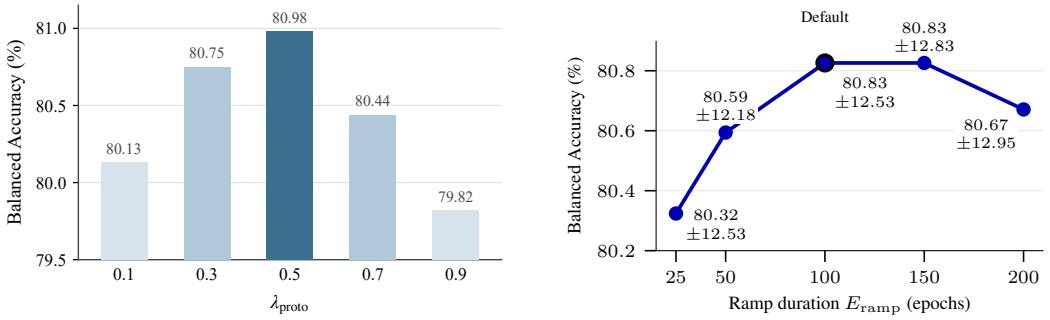


Figure 5: Source–sample transfer-factor distributions on BCI Competition IV-2a. Each panel corresponds to one target subject, and each violin shows the multi-granularity transfer factors of samples from one source subject. The dashed line at one denotes uniform source contribution, the gray column denotes the excluded target subject, and violin colors encode the mean transfer factor of the corresponding source subject.

We also examine the progressive ramp duration E_{ramp} on BCI-IV-2a (Fig. 6b).



(a) λ_{proto} sensitivity.

(b) Ramp duration sensitivity.

Figure 6: Parameter sensitivity on BCI-IV-2a. (a) Balanced accuracy when varying λ_{proto}, with $\lambda_{\text{cluster}} = 1 - \lambda_{\text{proto}}$. (b) Mean balanced accuracy when varying the progressive ramp duration E_{ramp}; annotations report mean $\pm$ subject-wise standard deviation across target subjects. The default setting is $E_{\text{ramp}} = 100$.

4 Conclusion

We presented Multi-Granularity Transfer (MGT), a general transfer framework for cross-subject EEG adaptation that models transferability across different source subjects and individual source samples. MGT combines coarse- and fine-grained transfer factors and progressively incorporates the resulting multi-granularity transfer factors during training to regulate source contributions. Experiments on seven EEG datasets spanning five paradigms show strong cross-subject decoding performance. Ablation studies support the roles of coarse-grained transfer, fine-grained transfer, and progressive transfer, while the transfer analyses provide qualitative and quantitative evidence that transferability varies across source subjects and individual source samples. MGT introduces additional computation through reference-space pretraining and transfer-factor estimation. Future work will investigate more efficient transferability estimation and extend MGT toward calibration-free cross-subject adaptation.

AI USE STATEMENT

Generative AI tools were used to assist with language editing and polishing, literature retrieval and discovery, and drafting and revising portions of the manuscript. Generative AI tools were not used for research ideation or execution, generating synthetic datasets, or proving mathematical claims. All AI-assisted text and literature-related outputs were reviewed and verified by the authors. The authors take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide implementation and experimental details to support reproducibility. Dataset preprocessing and evaluation protocols, training configurations, and backbone optimization details are provided in Appendix B. The sources of comparison results and the relationship between the original and updated GAT implementations are documented in Appendix C. Additional experimental results and subject-wise ablations are reported in Appendix D, and extended transfer analyses are provided in Appendix E.

REFERENCES

Kai Keng Ang, Zhengyang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE International Joint Conference on Neural Networks*, pp. 2390–2397, 2008. doi: 10.1109/IJCNN.2008.4634130.

Phairot Autthasan, Rattanaphon Chaisaen, Thapanun Sudhawiyangkul, Phurin Rangpong, Suktipol Kiatthaveephong, Nat Dilokthanakul, Gun Bhakdisongkhram, Huy Phan, Cuntai Guan, and Theerawit Wilaiprasitporn. MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification. *IEEE Transactions on Biomedical Engineering*, 69(6):2105–2118, 2022. doi: 10.1109/TBME.2021.3137184.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. doi: 10.1109/TPAMI.2016.2615921.

Weigang Cui, Yansong Xiang, Yifan Wang, Tao Yu, Xiao-Feng Liao, Bin Hu, and Yang Li. Deep multiview module adaption transfer network for subject-specific EEG recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):2917–2930, 2025. doi: 10.1109/TNNLS.2024.3350085.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pp. 2292–2300, 2013.

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.

He He and Dongrui Wu. Transfer learning for brain-computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2020. doi: 10.1109/TBME.2019.2913914.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024.

Magdiel Jiménez-Guarneros and Pilar Gómez-Gil. Standardization-refinement domain adaptation method for cross-subject EEG-based classification in imagined speech recognition. *Pattern Recognition Letters*, 141:54–60, 2021. doi: 10.1016/j.patrec.2020.11.013.

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

O-Yeon Kwon, Min-Ho Lee, Cuntai Guan, and Seong-Whan Lee. Subject-independent brain-computer interfaces based on deep convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(10):3839–3852, 2020. doi: 10.1109/TNNLS.2019.2946869.

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

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.

Shanglin Li, Shiwen Chu, Okan Koç, Yi Ding, Qibin Zhao, Motoaki Kawanabe, and Ziheng Chen. HEEGNet: Hyperbolic embeddings for EEG. In *The Fourteenth International Conference on Learning Representations*, 2026.

Chenyu Liu, Xinliang Zhou, Zhengri Zhu, Liming Zhai, Ziyu Jia, and Yang Liu. VBH-GNN: Variational bayesian heterogeneous graph neural networks for cross-subject emotion recognition. In *The Twelfth International Conference on Learning Representations*, 2024.

Wei Lu, Xiaobo Zhang, Lingnan Xia, Hua Ma, and Tien-Ping Tan. Domain adaptation spatial feature perception neural network for cross-subject EEG emotion recognition. *Frontiers in Human Neuroscience*, 18:1471634, 2024. doi: 10.3389/fnhum.2024.1471634.

Huan Luo, Na Lu, Xiaopeng Wang, and Xu Niu. Bayesian test-time adaptation via dirichlet feature projection and GMM-driven inference for motor imagery EEG decoding. In *The Fourteenth International Conference on Learning Representations*, 2026.

Jie Luo, Weigang Cui, Song Xu, Lina Wang, Huiling Chen, and Yang Li. A cross-scale transformer and triple-view attention based domain-rectified transfer learning for EEG classification in RSVP tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:672–683, 2024. doi: 10.1109/TNSRE.2024.3359191.

Chuong H. Nguyen, George K. Karavas, and Panagiotis Artemiadis. Inferring imagined speech using EEG signals: A new approach using riemannian manifold features. *Journal of Neural Engineering*, 15(1):016002, 2018. doi: 10.1088/1741-2552/aa8235.

Van-Chien Nguyen, Trung-Hieu Tran, Tuan-Kiet Doan, Quang Hung Pham, Ngoc-Son Vu, Duc Han Le, Huy Phan, Phi Le Nguyen, Nikola Simidjievski, Samuel Tardieu, and Van-Tam Nguyen. Balanced latent semantics and signal fidelity for EEG representation learning. In *1st ICLR Workshop on Time Series in the Age of Large Models*, 2026.

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.

D. Sculley. Web-scale k-means clustering. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 1177–1178, 2010. doi: 10.1145/1772690.1772862.

Yonghao Song, Qingqing Zheng, Qiong Wang, Xiaorong Gao, and Pheng-Ann Heng. Global adaptive transformer for cross-subject enhanced EEG classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:2767–2777, 2023. doi: 10.1109/TNSRE.2023.3285309.

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.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

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, 2024a. doi: 10.52202/079017-1239.

He Wang, Peiyin Chen, Meng Zhang, Jianbo Zhang, Xinlin Sun, Mengyu Li, Xiong Yang, and Zhongke Gao. EEG-based motor imagery recognition framework via multisubject dynamic transfer and iterative self-training. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (8):10698–10712, 2024b. doi: 10.1109/TNNLS.2023.3243339.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *The Thirteenth International Conference on Learning Representations*, 2025.

Fulin Wei, Xueyuan Xu, Xiuxing Li, and Xia Wu. BDAN-SPD: A brain decoding adversarial network guided by spatiotemporal pattern differences for cross-subject MI-BCI. *IEEE Transactions on Industrial Informatics*, 20(12):14321–14329, 2024. doi: 10.1109/TII.2024.3450010.

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.

Liuyin Yang, Qiang Sun, Ang Li, and Marc M. Van Hulle. Are EEG foundation models worth it? comparative evaluation with traditional decoders in diverse BCI tasks. In *The Fourteenth International Conference on Learning Representations*, 2026.

Wen Zhang and Dongrui Wu. Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(5):1117–1127, 2020. doi: 10.1109/TNSRE.2020.2985996.

He Zhao, Qingqing Zheng, Kai Ma, Huiqi Li, and Yefeng Zheng. Deep representation-based domain adaptation for nonstationary EEG classification. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):535–545, 2021. doi: 10.1109/TNNLS.2020.3010780.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015. doi: 10.1109/TAMD.2015.2431497.

Yangxuan Zhou, Sha Zhao, Jiquan Wang, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. BrainUICL: An unsupervised individual continual learning framework for EEG applications. In *The Thirteenth International Conference on Learning Representations*, 2025.

APPENDIX

A RELATED WORK

A common strategy for cross-domain EEG transfer is to reduce the discrepancy between source and target distributions. Euclidean Alignment transforms EEG covariances into a common reference space to reduce inter-subject variability (He & Wu, 2020), while DRDA learns domain-invariant representations for nonstationary EEG classification (Zhao et al., 2021). Optimal-transport-based domain adaptation similarly reduces distributional mismatch by explicitly coupling source and target samples (Courty et al., 2017). These approaches mainly improve transfer by aligning source and target distributions, but they do not explicitly distinguish whether different source subjects or individual source samples contribute equally to a particular target subject.

More selective transfer strategies have been explored by conditioning knowledge transfer on source–target relationships. Wang et al. (2024b) use dynamic transfer to match source and target samples with higher similarity in the latent space, thereby reducing the influence of less compatible source samples. Cui et al. (2025) further adapt module-wise knowledge sharing for each target subject according to source–target feature distributions. These methods move beyond uniform source utilization by introducing target-dependent transfer selection. However, their transfer decisions are formed from either local source–target similarity or target-specific model adaptation, rather than jointly characterizing the overall compatibility of each source subject and the variation in transferability among samples within that source.

Recent cross-subject EEG methods also improve generalization through stronger representation learning and adaptation mechanisms. GAT learns globally adaptive representations for cross-subject EEG classification (Song et al., 2023), HEEGNet introduces coarse-to-fine adaptation in a hyperbolic representation space (Li et al., 2026), and BTTA-DG performs test-time adaptation to handle cross-subject and cross-session distribution shifts (Luo et al., 2026). In contrast, MGT explicitly estimates transferability at two complementary granularities: a coarse-grained factor evaluates the compatibility between each source subject and the target, while a fine-grained factor further distinguishes individual source samples using discriminative and local structural information. The two factors are fused and progressively introduced during training to regulate multiple source-dependent objectives. Thus, MGT focuses on heterogeneous source contribution itself, rather than only reducing global domain discrepancy or adapting a representation model to the target distribution.

B IMPLEMENTATION DETAILS

B.1 DATASET PREPROCESSING AND EVALUATION PROTOCOLS

For all datasets, we follow a leave-one-subject-out cross-subject protocol. One subject is treated as the target and all remaining subjects form the source pool. The designated labeled target calibration partition is used for adaptation and transferability estimation, while a separate target partition is used for evaluation. Dataset-specific preprocessing and partitioning procedures are described below.

BCI Competition IV-2a. We use the 22-channel motor-imagery EEG data from the nine subjects. For each subject, the official training partition is used for source training or target calibration, and the corresponding evaluation partition is used for evaluation. Both partitions are processed using the same Chebyshev band-pass filtering procedure. Per-channel z-score statistics are estimated exclusively from the training partition and are subsequently applied to both the training and evaluation data. For each target subject, the training partitions of the remaining eight subjects constitute the source pool.

BCI Competition IV-2b. We use the three EEG channels and extract the interval from 3 to 7 s after each trial onset. Continuous EEG is band-pass filtered between 4 and 40 Hz using a sixth-order Chebyshev Type-II filter with 40-dB stop-band attenuation. Per-channel z-score statistics are estimated from each subject's training partition and are also used to normalize its evaluation partition. The official training partition is used for source training or target calibration, while the corresponding evaluation partition is used for evaluation.

OpenBMI. OpenBMI is processed using the MOABB motor-imagery pipeline. EEG signals are band-pass filtered from 4 to 40 Hz, cropped to the 0–4 s interval relative to the motor-imagery event, and resampled from 1000 Hz to 250 Hz. Session 1 is used for source training and target calibration, whereas Session 2 is used for target evaluation. Per-channel z-score statistics are estimated from Session 1 only and are subsequently applied to both sessions.

SEED. We use the 62-channel EEG recordings resampled to 200 Hz. Each recording is partitioned into non-overlapping 1000-sample windows, corresponding to 5 s of EEG. For each target subject, Sessions 1 and 2 are used as labeled target calibration data and Session 3 is used for evaluation. The source pool consists of Sessions 1 and 2 from the remaining 14 subjects.

NM000252. We use the 60-channel EEG representation sampled at 256 Hz. Each original 5-s trial is represented by three 2-s windows of 512 samples. The training/evaluation partition is defined at the original-trial level before window extraction so that windows derived from the same trial do not appear in different partitions. For each target subject, 160 labeled trials, corresponding to 480 windows, are used for calibration, and the remaining trials are used for evaluation. Per-channel normalization statistics are estimated using the training windows only.

RSVP. We use the provided preprocessed RSVP trials with eight EEG channels. For each subject, trials are divided into a fixed stratified 80% training partition and 20% evaluation partition while preserving the class proportions. Source subjects contribute only their training partitions, and the 80% partition of the target subject is used for calibration and transferability estimation. Per-channel z-score statistics are estimated from the training partition and then applied to the corresponding evaluation partition.

ERN. We use the prepared 56-channel ERN representation at 200 Hz. Following the session ordering of the prepared data, Sessions 1–4, containing 240 trials in total, are used for source training or target calibration, while Session 5, containing 100 trials, is used for target evaluation. Per-channel normalization statistics are estimated exclusively from the training partition and are then applied to the evaluation partition.

B.2 Training and Transfer Configuration

The same MGT configuration is used across datasets unless a dataset-specific input dimension is required. The reference model is trained before transfer-factor estimation and is subsequently frozen. The final model uses an independent initialization. Table 5 summarizes the principal optimization and transfer-related hyperparameters.

For entropy-regularized optimal transport, at most 64 representations are sampled from each compared set. The source-level temperature τ_s is determined separately for each target subject using the median source-level discrepancy. If the median becomes numerically degenerate, the mean discrepancy and then a unit value are used as successive fallbacks. The resulting source-level factors are normalized by their mean before being clipped to $[\alpha_{\min}, \alpha_{\max}]$.

For fine-grained transfer, source representations are clustered independently within each class. The number of clusters is $K_c = \lceil |I_c|/m \rceil$ with $m = 64$. Standard K-means is used for smaller classes, while MiniBatch K-means is used when a class contains at least 512 source samples. Prototype and cluster costs are combined with equal coefficients, i.e., $\lambda_{\text{proto}} = \lambda_{\text{cluster}} = 0.5$.

The combined fine-grained cost is normalized independently within each class using its median and median absolute deviation (MAD). If the MAD is numerically zero, the within-class standard deviation is used as a fallback, followed by a unit scale if necessary. The coarse- and fine-grained factors are then multiplied and normalized to unit mean within each class before being clipped to $[w_{\min}, w_{\max}]$. The resulting transfer factors remain fixed during final training; only their effective contribution changes through the progressive transfer mechanism during the first $E_{\text{ramp}} = 100$ epochs.

Optional source-subject preselection is disabled in the reported experiments. Consequently, all non-target subjects participate in transfer-factor estimation and final training, and their contributions

Table 5: Optimization and transfer-related hyperparameters used in MGT.

Hyperparameter	Value
Reference pretraining epochs	80
Learning rate	2×10^{-4}
Adam β_1, β_2	$0.5,\ 0.999$
Batch size	64
Progressive ramp E_{ramp}	100
Nominal cluster size m	64
Maximum OT samples per set	64
Sinkhorn regularization η	0.10
Sinkhorn iterations	100
$\alpha_{\min}, \alpha_{\max}$	$0.3,\ 2.0$
λ_{proto}	0.5
λ_{cluster}	0.5
$\beta_{\min}$	0.3
τ_t	1.0
$w_{\min}, w_{\max}$	$0.2,\ 2.0$
Numerical constant ϵ	10^{-8}
Center-regularization coefficient λ_{cen}	0.5
Gradient-penalty coefficient	10

are regulated by the estimated source- and sample-level transfer factors rather than by hard source selection.

B.3 GAT Backbone and Optimization Details

MGT is implemented on the updated GAT backbone used throughout our experiments. The feature extractor contains complementary temporal–spatial and spatial–temporal convolutional branches, followed by a projection to 40-dimensional token representations. The resulting sequence is processed by a multi-layer Transformer containing self-attention and cross-attention operations. The classification head operates on the final Transformer representation, while a domain discriminator is jointly optimized to reduce source–target distribution mismatch.

The reference model and final model use the same backbone architecture but are trained independently. During reference-space pretraining, all source contributions are treated uniformly. After 80 epochs, the reference encoder is frozen and used only to obtain the representations required for coarse- and fine-grained transfer-factor estimation. The final model is then initialized independently and trained using the progressively introduced transfer factors.

The inherited GAT objective contains source classification, source–target alignment, and class-structure regularization terms. MGT does not replace these objectives. Instead, for each source-dependent term, the ordinary source mini-batch average is replaced by the transfer-aware aggregation introduced in Section 2.3. Target-dependent terms remain unweighted. The center-regularization coefficient is set to 0.5, and the gradient-penalty coefficient is set to 10.

We also retain the target-side segmentation-and-reconstruction (S&R) augmentation used by the GAT implementation. A synthesized target example is formed by recombining eight temporal segments drawn from labeled target calibration examples of the same class, and approximately half of the target examples participating in a training step are replaced by reconstructed examples. This augmentation is independent of transfer-factor estimation: MGT modifies the contribution of source knowledge, whereas S&R operates only on the labeled target calibration data.

At inference time, neither the reference model nor transfer-factor estimation is required. Source EEG data and target calibration labels are also no longer used. Prediction therefore requires only the final representation model and classifier.

C Comparison Methods and GAT Implementations

C.1 Sources of Comparison Results

For the motor-imagery benchmarks, the compared methods include FBCSP (Ang et al., 2008), EEGNet (Lawhern et al., 2018), ShallowConvNet (Schirrmeister et al., 2017), MIN2Net (Autthasan et al., 2022), Spectral-Spatial CNN (Kwon et al., 2020), DRDA (Zhao et al., 2021), BDAN-SPD (Wei et al., 2024), and GAT (Song et al., 2023). The SEED and NM000252 comparison values follow Lu et al. (2024) and Jiménez-Guarneros & Gómez-Gil (2021), respectively. For RSVP, the compared transfer methods include Euclidean Alignment (He & Wu, 2020), MEKT-R (Zhang & Wu, 2020), DANN (Ganin et al., 2016), and CST-TVA-DRTL (Luo et al., 2024). For ERN, the compared pretrained EEG models include BENDR (Kostas et al., 2021), BIOT (Yang et al., 2023), LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024a), CBraMod (Wang et al., 2025), and STELAR (Nguyen et al., 2026).

Unless otherwise stated, these comparison results are literature-reported. GAT (official updated) and MGT are evaluated under the same implementation and evaluation setting.

C.2 Original and Updated GAT Implementations

The updated GAT backbone used in our experiments is described in Appendix B.3. Here, we clarify why the originally reported GAT result and GAT (official updated) are listed separately in Table 2. The original study (Song et al., 2023) reports a parallel convolutional feature extractor followed by a separate global adaptor, domain discriminator, and classifier. The subsequently released official v0.72 changes the feature extractor, adaptation module, Transformer structure, and classifier, as summarized in Table 6.

Table 6: Main differences between the originally reported GAT and the subsequently released official v0.72 implementation used in our experiments.

Component	Original GAT	Official v0.72
Feature extractor	Parallel temporal and spatial convolutional branches	Temporal–spatial and spatial–temporal branches followed by token projection
Adaptation module	Separate attention-based global adaptor	Encoder–decoder Transformer with integrated self- and cross-attention
Transformer structure	Single global adaptation stage	Three stacked Transformer blocks with 40-dimensional tokens
Classifier	Two fully connected layers	Flattened Transformer representation followed by a multi-layer MLP

The two GAT results are reported separately because the implementations differ. MGT uses the official v0.72 backbone, making GAT (official updated) the same-backbone comparison under the shared evaluation setting.

D Additional Experimental Results

D.1 Results on BCI Competition IV-2b

Table 7 reports the results on BCI Competition IV-2b, the seventh dataset included in our evaluation. MGT achieves a balanced accuracy of 85.33% and a Cohen's Kappa of 0.7065. Under the same implementation and evaluation setting, GAT (official updated) obtains 84.78% balanced accuracy and 0.6955 Kappa. MGT therefore improves balanced accuracy by 0.55 percentage points and Kappa by 0.0110, while reducing the subject-wise standard deviation from 13.22% to 12.37%.

Table 7: Cross-subject decoding results on BCI Competition IV-2b. Std. denotes subject-wise standard deviation.

Method	Balanced Acc. (%) ↑	Kappa ↑	Std. ↓
RA-MDRM	62.41	0.5102	11.08
EA-CSP-LDA	62.50	0.5217	11.09
ssCSP	75.09	0.5018	13.97
SSMM	78.00	0.5600	14.41
ShallowConvNet	79.37	0.5875	17.26
EEGNet	82.37	0.6507	**10.15**
MIN2Net	84.60	0.6921	11.46
GAT (official updated)	84.78	0.6955	13.22
MGT (Ours)	**85.33**	**0.7065**	12.37

These results extend the evaluation to a two-class motor-imagery benchmark with only three EEG channels. Compared with GAT (official updated) under the same implementation and evaluation setting, MGT improves both balanced accuracy and Kappa.

D.2 Subject-Wise Ablation Results on BCI Competition IV-2a

To complement the aggregate ablation results in the main text, we report subject-wise results for all nine target subjects on BCI Competition IV-2a. Tables 8 and 9 report balanced accuracy and Cohen's Kappa, respectively.

Table 8: Subject-wise balanced accuracy (%) for the BCI Competition IV-2a ablation study.

Variant	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
w/o Coarse-Grained Transfer	92.36	64.24	93.40	80.90	68.06	67.01	94.79	88.54	86.81	81.79
w/o Fine-Grained Transfer	92.36	64.58	92.71	82.29	68.40	67.71	96.18	87.15	87.85	82.14
w/o Progressive Transfer	91.67	61.81	93.06	84.38	68.06	67.01	93.75	88.54	87.85	81.79
Full MGT	92.71	64.93	93.06	82.99	73.26	71.18	93.75	88.54	88.19	**83.18**

Table 9: Subject-wise Cohen's Kappa for the BCI Competition IV-2a ablation study.

Variant	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
w/o Coarse-Grained Transfer	0.8981	0.5231	0.9120	0.7454	0.5741	0.5602	0.9306	0.8472	0.8241	0.7572
w/o Fine-Grained Transfer	0.8981	0.5278	0.9028	0.7639	0.5787	0.5694	0.9491	0.8287	0.8380	0.7618
w/o Progressive Transfer	0.8889	0.4907	0.9074	0.7917	0.5741	0.5602	0.9167	0.8472	0.8380	0.7572
Full MGT	0.9028	0.5324	0.9074	0.7731	0.6435	0.6157	0.9167	0.8472	0.8426	**0.7757**

The effect of each component varies across target subjects. Individual ablated variants can outperform Full MGT on some targets, whereas Full MGT achieves the highest subject-averaged balanced accuracy and Kappa. The differences are particularly pronounced for S5 and S6, indicating that the contribution of the transfer components depends on the target subject.

D.3 Subject-Wise Ablation Results on SEED

We further report subject-wise ablation results for all 15 target subjects on SEED. Tables 10 and 11 provide balanced accuracy and Cohen's Kappa, respectively.

Table 10: Subject-wise balanced accuracy (%) for the SEED ablation study.

Variant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	Mean
w/o Coarse-Grained Transfer	75.69	72.72	73.72	60.70	35.36	85.39	85.80	83.74	84.71	80.32	99.85	66.71	78.29	77.52	98.79	77.29
w/o Fine-Grained Transfer	73.13	75.96	74.43	74.45	32.53	90.81	93.11	82.76	90.69	63.61	97.85	66.35	73.39	76.64	98.70	77.63
w/o Progressive Transfer	70.24	72.57	71.69	57.54	37.68	86.24	89.66	85.18	80.32	81.26	100.00	66.03	80.17	76.08	100.00	76.98
Full MGT	73.13	74.10	84.71	59.10	56.07	93.64	91.51	87.92	83.59	87.89	100.00	67.38	82.12	83.50	96.99	**81.44**

Table 11: Subject-wise Cohen's Kappa for the SEED ablation study.

Variant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	Mean
w/o Coarse-Grained Transfer	0.6365	0.5920	0.6058	0.4121	0.0411	0.7804	0.7869	0.7567	0.7706	0.7042	0.9978	0.4980	0.6736	0.6619	0.9819	0.6600
w/o Fine-Grained Transfer	0.5976	0.6399	0.6160	0.6168	0.0000	0.8620	0.8966	0.7421	0.8604	0.4549	0.9677	0.4913	0.5998	0.6486	0.9805	0.6650
w/o Progressive Transfer	0.5553	0.5896	0.5749	0.3634	0.0750	0.7931	0.8448	0.7782	0.7053	0.7179	1.0000	0.4869	0.7020	0.6402	1.0000	0.6551
Full MGT	0.5982	0.6128	0.7707	0.3867	0.3445	0.9045	0.8727	0.8191	0.7538	0.8181	1.0000	0.5084	0.7313	0.7518	0.9549	**0.7218**

The results again show substantial variation across target subjects. Different ablated variants can perform better for individual targets, but Full MGT achieves the highest subject-averaged balanced accuracy and Kappa over the 15 subjects. This variation further indicates that the effect of the transfer components is target dependent.

D.4 Subject-Wise Ablation Results on NM000252

Finally, we report subject-wise ablation results for all six target subjects in NM000252. Tables 12 and 13 report balanced accuracy and Cohen's Kappa, respectively.

Table 12: Subject-wise balanced accuracy (%) for the NM000252 ablation study.

Variant	S1	S2	S3	S4	S5	S6	Mean
w/o Coarse-Grained Transfer	57.50	74.17	65.00	59.17	63.33	56.67	62.64
w/o Fine-Grained Transfer	56.67	73.33	60.00	58.33	60.83	55.83	60.83
w/o Progressive Transfer	55.83	74.17	60.83	58.33	64.17	58.33	61.94
Full MGT	60.00	75.00	62.50	58.33	64.17	56.67	**62.78**

Table 13: Subject-wise Cohen's Kappa for the NM000252 ablation study.

Variant	S1	S2	S3	S4	S5	S6	Mean
w/o Coarse-Grained Transfer	0.1500	0.4833	0.3000	0.1833	0.2667	0.1333	0.2528
w/o Fine-Grained Transfer	0.1333	0.4667	0.2000	0.1667	0.2167	0.1167	0.2167
w/o Progressive Transfer	0.1167	0.4833	0.2167	0.1667	0.2833	0.1667	0.2389
Full MGT	0.2000	0.5000	0.2500	0.1667	0.2833	0.1333	**0.2556**

The subject-wise results show that no individual component has an identical effect across all targets. Full MGT achieves the highest subject-averaged balanced accuracy and Kappa, while individual ablated variants can match or exceed it for some subjects. These results are consistent with the heterogeneous transferability observed across target subjects.

E Additional Analysis

E.1 Extended Sample-Level Transfer Patterns

The main text visualizes sample-level transfer patterns for representative target subjects on BCI Competition IV-2a. Here, we extend the visualization to all nine target subjects. Following the same setting as in the main text, source samples with fused transfer factors at or below the within-class 30th percentile are treated as low-factor samples, while those at or above the 70th percentile are treated as high-factor samples. Source samples with intermediate transfer factors are omitted from the visualization.

Figure 7 shows substantial variation in the spatial relationship between source and target samples across target subjects. High-factor source samples frequently occupy regions that are more closely associated with the target calibration samples, whereas low-factor samples are more dispersed in many subjects. At the same time, the spatial patterns are not identical across targets, indicating that sample-level transferability depends on the target subject rather than following a fixed pattern across subjects. These extended visualizations are consistent with the representative sample-level transfer patterns reported in the main text.

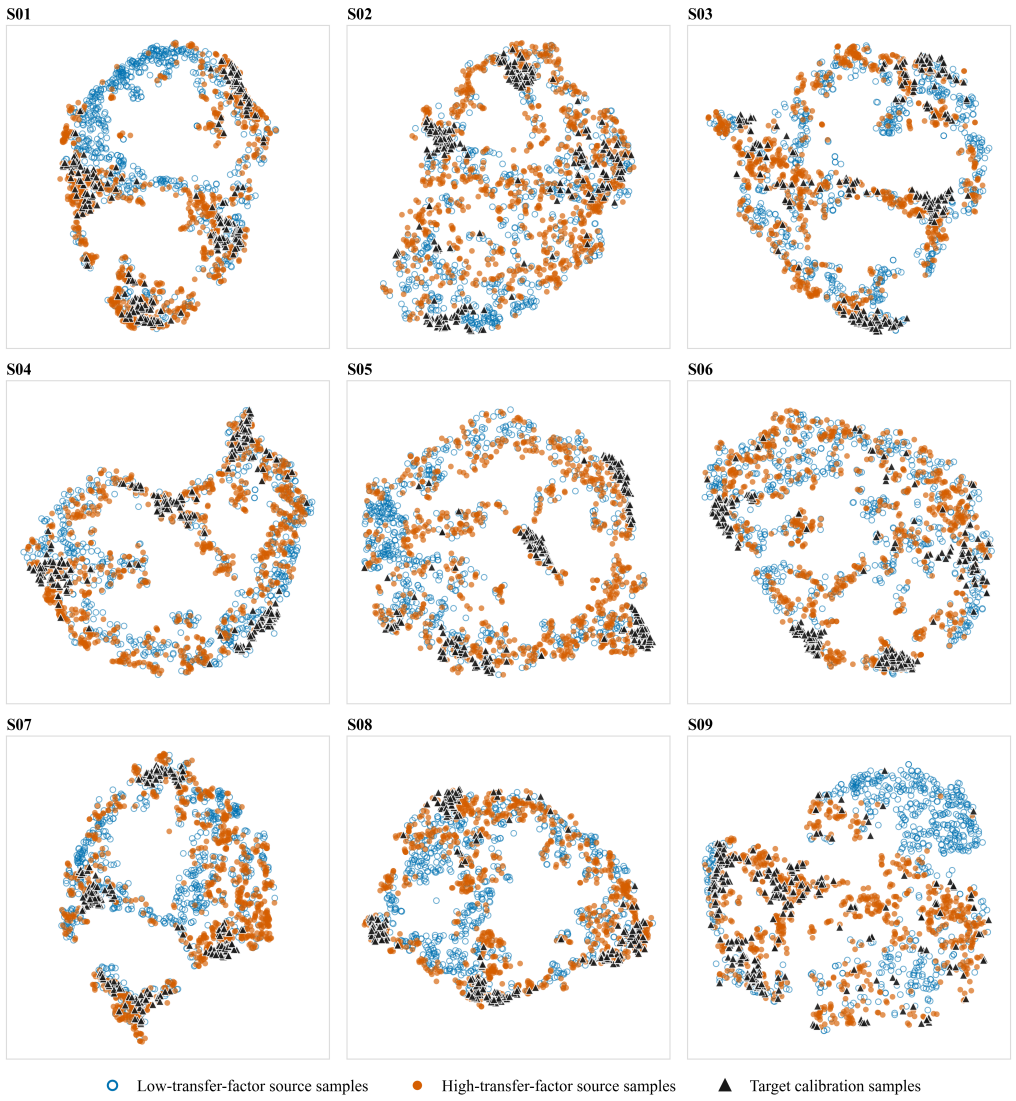


Figure 7: Extended sample-level transfer patterns on BCI Competition IV-2a. t-SNE embeddings of mean-pooled 40-dimensional final-model representations are shown for all nine target subjects. Blue hollow circles and orange filled circles denote source samples with fused transfer factors at or below the within-class 30th percentile and at or above the 70th percentile, respectively; black triangles denote target calibration samples. Source samples with intermediate transfer factors are omitted from display.

E.2 Subject-Wise Confusion Analysis

The aggregate confusion matrix in the main text summarizes class-wise decoding behavior across all target subjects. To further examine cross-subject variation, Figure 8 reports the confusion matrix separately for each of the nine target subjects on BCI Competition IV-2a.

The subject-wise matrices reveal substantial variation in decoding behavior across targets. S1, S3, S7, and S8 show relatively strong discrimination across all four motor-imagery classes, whereas S2, S5, and S6 exhibit more pronounced class confusion. For S2, the largest difficulty occurs for left-hand and right-hand imagery, with recalls of 44.44% and 55.56%, respectively. For S5 and S6, tongue imagery is more difficult, with recalls of 55.56% and 56.94%.

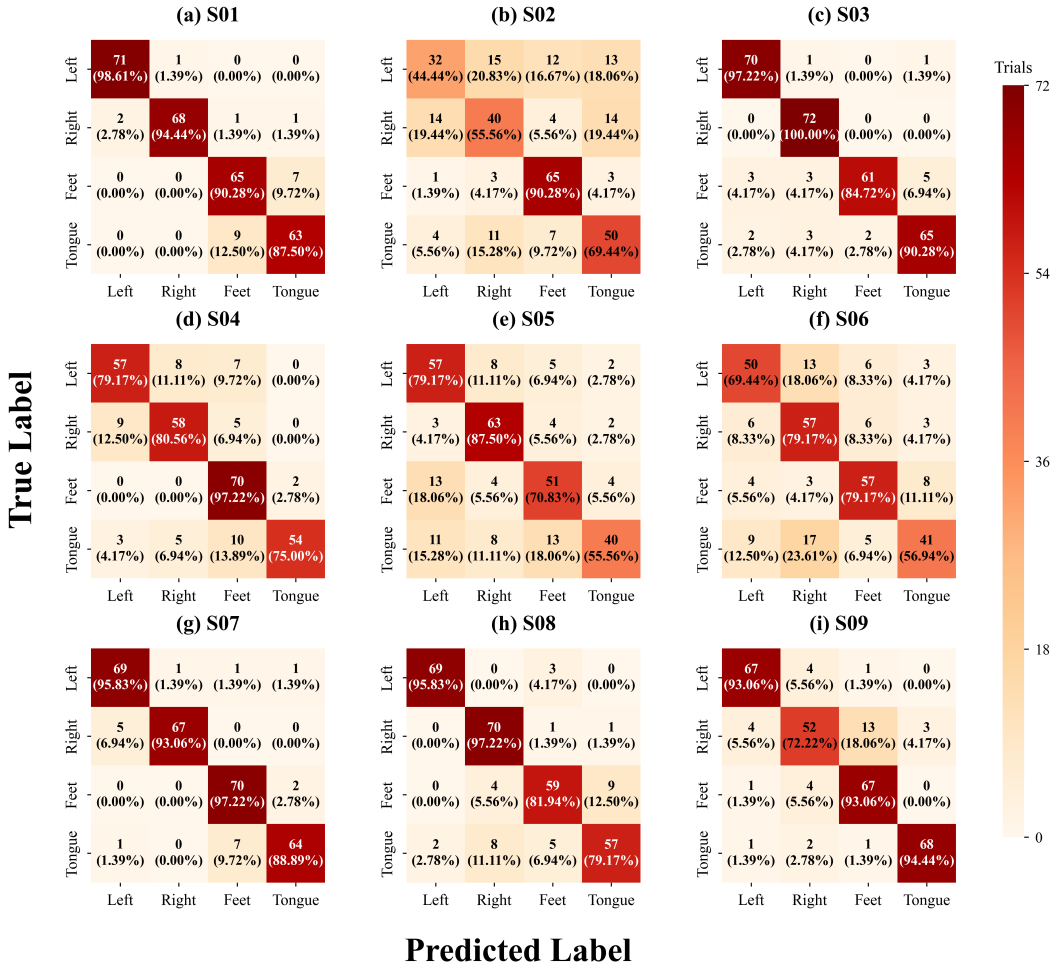


Figure 8: Subject-wise confusion matrices of MGT on BCI Competition IV-2a. Each panel corresponds to one target subject. Rows denote true labels and columns denote predicted labels. Each cell reports the number of trials together with the corresponding row-normalized percentage.

The error patterns are also target dependent. For example, S9 shows lower recall for right-hand imagery than for its other classes, while S4 retains high recall for feet imagery but exhibits more confusion for the remaining classes. These subject-specific differences complement the aggregate confusion analysis in the main text and further illustrate the heterogeneity of cross-subject EEG decoding.