GeoLPE: Geometry-Aware Latent Prediction for EEG Foundation Modeling

Anonymous authors
Paper under double-blind review

Abstract

Electroencephalography (EEG) foundation models learn reusable representations across datasets and tasks, yet heterogeneous electrode montages complicate spatial transfer, while waveform reconstruction may emphasize acquisition-specific signal details. We introduce GeoLPE, a geometry-aware latent-prediction framework for EEG foundation modeling. GeoLPE combines continuous absolute spatiotemporal encoding with explicit relative inter-electrode geometry and learns target representations through structured joint-embedding prediction. We pretrain GeoLPE on 97 public EEG dataset entries and evaluate it on eight EEG datasets spanning motor imagery, event-related potentials (ERP), code-modulated visual evoked potentials (c-VEP), and cognitive workload. Among the evaluated models, GeoLPE achieves the highest cross-dataset mean balanced accuracy under all four evaluation protocols, covering cross-subject and within-subject settings with both full fine-tuning and linear probing. It reaches 69.70% under cross-subject full fine-tuning, compared with 67.54% for the strongest baseline average. The pretrained encoder is adapted to electrocorticography (ECoG) finger-trajectory prediction, achieving a mean Pearson correlation of 0.566. Ablations support the contributions of latent prediction, auxiliary objectives, and continuous geometric modeling, while representation analyses demonstrate improved transferability of pretrained representations across participants and train–test partitions. These results provide a foundation for scalable, efficient, and generalizable EEG modeling.

1 Introduction

Electroencephalography (EEG) is widely used to detect abnormal brain activity, assess brain states, and decode behavior-related neural activity (Kostas et al., 2021; Jiang et al., 2024; Wang et al., 2024; 2025). These applications span diverse neural processes and temporal scales, with data collected under heterogeneous experimental paradigms and acquisition protocols.

EEG foundation models learn shared encoders through multi-source pretraining to support diverse downstream tasks. Cross-source encoder reuse requires addressing two sources of recording variability. Electrode montages determine spatial relationships among channels, while acquisition systems and non-neural interference affect signal statistics. Spatial modeling and pretraining objectives are therefore central to encoder transferability.

EEG signals are sampled by discrete electrodes across the scalp. Variations in electrode count, position, spacing, or coverage make channel indices unreliable spatial references. Coordinate-based representations can provide a common spatial reference across electrode montages (El Ouahidi et al., 2025; Wang et al., 2026). Absolute electrode positions and relative inter-electrode geometry play complementary roles. Absolute positions identify recording locations, while relative geometry guides information aggregation across electrodes. This distinction motivates joint modeling of absolute positions and relative geometry in a unified coordinate system.

Spatial modeling captures electrode geometry, whereas the pretraining objective shapes the signal representations learned by the encoder. Many EEG foundation models incorporate waveform reconstruction into their pretraining objectives (Wang et al., 2024; 2025). However, waveform-level losses are sensitive to noise and amplitude variations, potentially emphasizing signal details unrelated to downstream tasks. EEG2Rep explores an alternative by predicting latent representations rather than

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.

reconstructing waveforms (Mohammadi Foumani et al., 2024).This motivates latent prediction objectives that exploit structured spatiotemporal context without relying primarily on waveform reconstruction.

To address these challenges, we introduce GeoLPE, a geometry-aware latent-prediction framework for EEG foundation modeling. GeoLPE encodes absolute spatiotemporal positions and relative inter-electrode geometry in a unified continuous coordinate system. A single encoder supports variable electrode counts, diverse spatial layouts, and variable-length recordings, enabling joint pretraining across heterogeneous EEG datasets. During pretraining, GeoLPE predicts target latent representations from partial EEG observations using a structured joint-embedding objective. This objective places the primary pretraining supervision in latent space rather than waveform space.

We further investigate cross-modal transfer from scalp EEG to electrocorticography (ECoG). The two recording modalities differ in measurement geometry and spectral characteristics. During ECoG adaptation, the pretrained GeoLPE encoder serves as the low-frequency pathway, while a complementary pathway models high-gamma activity (Miller et al., 2007). Experiments on eight scalp EEG datasets and ECoG finger-trajectory prediction evaluate the reuse of the same pretrained encoder across datasets, tasks, and recording modalities.

Our contributions are threefold:

- GeoLPE jointly encodes absolute spatiotemporal positions and relative inter-electrode geometry in a unified continuous coordinate system, supporting heterogeneous electrode montages and variable-length recordings with a single encoder.
- GeoLPE adopts a structured joint-embedding predictive architecture (JEPA) for EEG pretraining, placing its primary supervision in latent space rather than waveform space.
- Across eight EEG datasets, GeoLPE achieves the best average performance among evaluated methods under cross-subject and within-subject evaluation with both full fine-tuning and linear probing. Its pretrained encoder also adapts to ECoG finger-trajectory prediction.

2 Related Work

EEG spatiotemporal positional encoding As pretraining scales across multiple datasets, EEG models must accommodate heterogeneity in electrode montage. REVE generates continuous four-dimensional positional encodings from three-dimensional electrode coordinates and temporal positions (El Ouahidi et al., 2025). BSR uses electrode coordinates to render EEG spectra as spatially structured videos (Wang et al., 2026). Despite these advances, existing positional modeling still struggles to jointly encode absolute electrode positions and inter-electrode directions and distances within a unified continuous coordinate system. When electrode count, spacing, or coverage varies, learned spatial relationships become less reusable.

Reconstruction Objectives in Pretraining The central distinction among pretraining objectives lies in the representation space for prediction error computation. MAE computes this error in observation space by reconstructing masked input pixels (He et al., 2022). I-JEPA instead predicts encoded image representations (Assran et al., 2023). V-JEPA extends latent prediction to video representations (Bardes et al., 2024). EEG foundation models likewise define supervision over distinct targets. CBraMod reconstructs waveform patches (Wang et al., 2025). LaBraM predicts discrete neural codes (Jiang et al., 2024). EEGPT jointly optimizes waveform recovery and alignment with representations from a momentum encoder (Wang et al., 2024). These objectives supervise continuous signals, discrete codes, or a hybrid of signals and representations. However, few EEG foundation models define their primary supervision entirely within the latent space of a target encoder.

ECoG Decoding and Cross-Modal Transfer Early ECoG decoding relied on task-specific networks, such as CNN–LSTM, which combines convolutional feature extraction with temporal regression to predict finger trajectories (Xie et al., 2018). For cross-subject and cross-modal transfer, HTNet extracts spectral power in data-driven frequency bands using the Hilbert transform and projects electrode-level features onto predefined brain regions (Peterson et al., 2021). Trained on multi-subject ECoG data, it was evaluated on unseen ECoG participants and scalp EEG participants. These

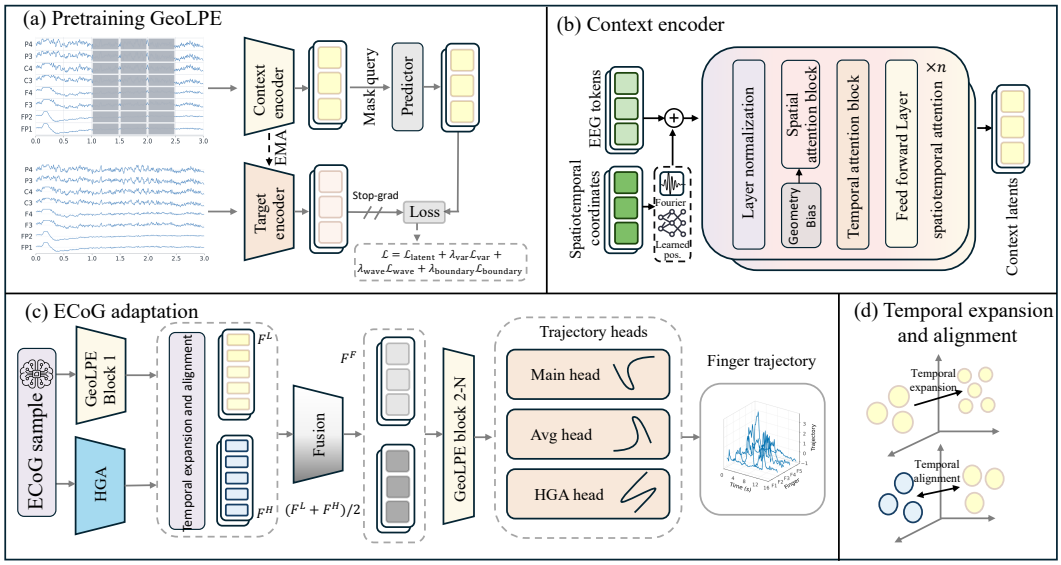

Figure 1: Overview of GeoLPE. (a) Structured latent-prediction pretraining predicts target-encoder representations from visible EEG context. (b) The context encoder combines continuous spatiotemporal positional encoding, geometry-conditioned spatial attention, and factorized spatiotemporal modeling. (c) For ECoG adaptation, pretrained low-frequency representations are aligned and fused with an HGA pathway for finger-trajectory prediction. (d) Low-frequency representations are temporally expanded to the HGA grid and aligned with HGA representations before fusion.

methods primarily learn directly from ECoG data, leaving spatiotemporal priors learned through large-scale scalp EEG pretraining underutilized in ECoG modeling.

3 METHOD

GeoLPE is the pretrained EEG foundation model obtained by combining a geometry-conditioned spatiotemporal encoder with structured latent-prediction pretraining. After pretraining, the context encoder is retained as the GeoLPE encoder for downstream scalp EEG tasks and ECoG adaptation. Figure 1 presents the complete framework. We represent a multichannel EEG recording as $X \in \mathbb{R}^{C\times T\times L}$, where C is the number of electrodes, T the number of temporal patches per channel, and L the number of samples per patch. The t-th waveform patch from electrode i is denoted by $x_{i,t} \in \mathbb{R}^L$. Electrode coordinates are given by $\mathcal{P} = (p_1, \ldots, p_C)$, where $p_i \in \mathbb{R}^3$ specifies the position of electrode i in a common coordinate system.

3.1 GEOMETRY-CONDITIONED SPATIOTEMPORAL ENCODER

Electrode montages vary across datasets in channel count, spacing, and coverage. To accommodate these variations, we combine factorized spatiotemporal attention with continuous geometry encoding. Each encoder layer applies spatial attention at each temporal position to model inter-electrode dependencies, followed by temporal attention within each electrode to capture temporal dynamics.

For input EEG patches, we adopt REVE's continuous spatiotemporal positional encoding over four-dimensional coordinates (El Ouahidi et al., 2025). Given the spatiotemporal coordinate $s_{i,t} = [p_i; t] \in \mathbb{R}^4$, a fixed Fourier mapping ϕ_{fix} and a learnable coordinate mapping ϕ_{learn} jointly produce the positional embedding:

$$e_{i,t} = \text{LayerNorm}(\phi_{\text{fix}}(s_{i,t}) + \phi_{\text{learn}}(s_{i,t}))\,.$$

The encoder input at position (i, t) is $f_{\text{sig}}(x_{i,t}) + e_{i,t}$, where $f_{\text{sig}} : \mathbb{R}^L \to \mathbb{R}^D$ is a linear projection shared across electrodes and temporal positions.

Beyond this input-level positional encoding, we explicitly model continuous relative electrode geometry within spatial attention.

Continuous relative geometry. For each spatial attention layer and head, let $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ denote the queries, keys, and values at a fixed temporal position. We define a spatial rotary operator $\mathcal{R}_{\mathcal{P}}$ determined by the electrode coordinates. The operator first centers these coordinates and uses a Cayley transform to parameterize a learnable orthogonal feature basis shared by queries and keys. Within this basis, paired feature dimensions are rotated by angles given by inner products between the centered coordinates and learnable 3D spatial frequencies (Su et al., 2024; Heo et al., 2024; Schenck et al., 2025). Sharing the basis and frequency parameters across queries, keys, and electrode positions ensures that the dependence introduced by the rotations is determined by relative electrode displacements.

We also construct a geometric bias matrix $B_{\mathcal{P}} \in \mathbb{R}^{C\times C}$ for the current montage. Each entry $(B_{\mathcal{P}})_{ij}$ is obtained by linearly mapping the unit direction vector from key electrode j to query electrode i together with their normalized distance. Distances are normalized by the maximum pairwise electrode distance in the set. With head dimension d_h, spatial attention is:

$$\mathrm{SA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}; \mathcal{P}) = \mathrm{softmax}\left(\frac{\mathcal{R}_{\mathcal{P}}(\mathbf{Q})\mathcal{R}_{\mathcal{P}}(\mathbf{K})^{\top}}{\sqrt{d_h}} + B_{\mathcal{P}}\right)\mathbf{V}. \tag{1}$$

The rotary operator modulates query–key matching, while the geometric bias adjusts pairwise attention logits. The continuous spatiotemporal positional encoding is added to EEG patch inputs, whereas the rotary operator and geometric bias are applied only within spatial attention. Appendix 9 gives the complete per-head computation.

3.2 GeoLPE Pretraining

GeoLPE is trained under the JEPA paradigm to predict latent representations of masked regions from visible electrode–time patches (Assran et al., 2023). The context and target encoders share the architecture defined in Section 3.1. Structured masking defines the target spatiotemporal regions. The predictor estimates representations of an unperturbed reference view using only visible patches from a perturbed context view.

Views and structured masking. To improve robustness to small electrode-position deviations and recording-level amplitude variations, we perturb only the context view. The context encoder processes visible patches from the perturbed view $(\widetilde{X}, \widetilde{\mathcal{P}})$, with $\widetilde{X} = \gamma X$ and $\widetilde{p}_i = p_i + \xi_i$. The coordinate perturbation ξ_i is shared across all temporal patches of electrode i. A positive gain $\gamma > 0$ is shared across the recording, preserving relative amplitudes across electrodes and time. The perturbation distributions are specified in Appendix 10.

To model both temporal and inter-electrode dependencies, we select temporal-band masking or spatial electrode-cluster masking with equal probability for each recording, targeting a masking ratio of 50%. We denote the target positions by $\mathcal{M}$ and the remaining visible positions by $\mathcal{V}$. Temporal masking covers a contiguous time interval across all electrodes. Spatial masking selects neighboring electrodes based on the reference 3D coordinates and covers all their temporal segments.

Latent prediction from visible context. The context encoder E_θ extracts representations from visible positions $\mathcal{V}$. The target encoder $E_{\bar{\theta}}$ processes the complete reference view. Its parameters are updated only by an exponential moving average of the context encoder parameters. We apply per-patch layer normalization and stop-gradient to its final-layer outputs to obtain target representations $z_{i,t}$. These representations serve as prediction targets.

The latent predictor G_ψ receives context representations at visible positions. At masked positions, it receives the sum of a shared mask query q_{mask} and the continuous spatiotemporal positional embedding. It outputs $\widehat{z}_{i,t}$ at all electrode–time positions. Outputs at target positions serve the main prediction objective, while outputs at visible positions support the mask-boundary objective. The predictor's global attention heads attend to all electrode–time positions. Its local heads attend only to temporal neighbors within the same electrode and spatial neighbors at the same temporal position. Spatial neighborhoods are determined from the reference electrode coordinates.

Training objective. We normalize the predictor outputs at target positions as: $\overline{z}_{i,t} = \mathrm{LayerNorm}(\widehat{z}_{i,t})$ and define the main latent prediction loss as:

$$\mathcal{L}_{\mathrm{latent}} = \frac{1}{|\mathcal{M}|} \sum_{(i,t)\in\mathcal{M}} \mathrm{SmoothL1}(\overline{z}_{i,t}, z_{i,t}).$$

This objective trains GeoLPE to infer the target encoder's full-view latent representations at the corresponding positions from visible context. We supplement it with three auxiliary objectives. The variance-floor loss $\mathcal{L}_{\mathrm{var}}$ computes the standard deviation of each predicted feature dimension across target positions and penalizes dimensions whose variation falls below a fixed margin. For waveform recovery, visible context representations are aggregated by attention pooling and combined with the positional embedding of each masked target. A training-only decoder reconstructs the corresponding reference waveform patch, and $\mathcal{L}_{\mathrm{wave}}$ is computed as the ℓ_1 reconstruction error over masked positions. The mask-boundary loss $\mathcal{L}_{\mathrm{boundary}}$ aligns predictor and target representations at visible positions using Smooth L1 distance, with larger weights assigned to positions closer to the nearest masked region in normalized spatiotemporal coordinates. The total objective is:

$$\mathcal{L} = \mathcal{L}_{\mathrm{latent}} + \lambda_{\mathrm{var}}\mathcal{L}_{\mathrm{var}} + \lambda_{\mathrm{wave}}\mathcal{L}_{\mathrm{wave}} + \lambda_{\mathrm{boundary}}(\tau)\mathcal{L}_{\mathrm{boundary}},$$

where τ denotes normalized training progress and controls the weight of the boundary objective. The predictor and auxiliary modules are used only during pretraining. We retain E_θ as the GeoLPE encoder for downstream tasks. Appendix 10 provides the complete loss definitions, weights, EMA schedule, and pretraining procedure.

3.3 Adaptation from Scalp EEG to ECoG

For ECoG adaptation, we initialize the low-frequency pathway with the pretrained GeoLPE encoder E_θ to reuse spatiotemporal representations learned from scalp EEG. The waveform representation and continuous geometric encoding defined in Section 3.1 also apply to ECoG recordings.

ECoG captures high-gamma activity (HGA) associated with local cortical population activity. Prior work has reported coupling between θ phase and high-gamma power, indicating a cross-frequency relationship between low-frequency rhythms and HGA (Canolty et al., 2006). Motivated by this relationship, we introduce an HGA pathway alongside the pretrained low-frequency pathway. The two pathways encode low-frequency waveforms and HGA amplitude envelopes, respectively, and fuse the resulting representations after electrode–time alignment.

Low-frequency and high-gamma encoding. The two pathways use different patch durations for each ECoG window. The low-frequency pathway retains the patch duration used during GeoLPE pretraining, while the HGA pathway divides amplitude envelopes into shorter patches and applies a temporal Transformer with parameters shared across electrodes to model within-electrode dynamics. This yields complementary temporal granularities, with a finer temporal grid in the HGA pathway. Signal filtering and HGA envelope extraction are detailed in Appendix 7.3.

Both pathways use the continuous spatiotemporal positional encoding defined in Section 3.1 with the same electrode order and coordinates. For electrode i, we concatenate its common-coordinate position p_i with a layout-normalized coordinate $\widehat{p}_i$, obtained by centering at the electrode centroid and scaling by the maximum electrode-to-centroid distance. A shared mapping produces the coordinate residual $c_i = f_{\mathrm{coord}}([p_i; \widehat{p}_i]) \in \mathbb{R}^D$, which is added to every patch embedding from electrode i in both pathways.

Dual-scale fusion and trajectory prediction. After the first GeoLPE encoder block, low-frequency representations are repeated to the corresponding HGA temporal positions to form F^{L}, aligning them with the HGA temporal Transformer output F^{H}. A gated fusion module then combines F^{L} and F^{H} to produce the fused representation F^{F}.

The main and average-fusion branches pass F^{F} and $(F^{\mathrm{L}} + F^{\mathrm{H}})/2$, respectively, through the remaining GeoLPE encoder blocks, while the HGA branch operates directly on F^{H}. Each branch uses a separate prediction head with mean and attention pooling to aggregate electrode–time representations and predict the five finger-trajectory values at a given decoding time. These branches preserve complementary predictions from the fused, HGA-specific, and simple average-fusion representations.

We additionally introduce a cross-frequency latent prediction objective over matched temporal support. HGA representations within each coarse low-frequency interval are averaged and passed through a training-only predictor to estimate the corresponding low-frequency latent representation. The mean cosine distance to the stop-gradient low-frequency target defines $\mathcal{L}_{\mathrm{cf}}$.

For ECoG window b, the main branch predicts $\widehat{y}_b^{\mathrm{main}} \in \mathbb{R}^K$, where $K = 5$ corresponds to the five fingers. The target $y_b \in \mathbb{R}^K$ contains the trajectory values at the decoding time and is normalized using training-set statistics. For a mini-batch of B_{E} windows, the main regression loss is:

$$\mathcal{L}_{\mathrm{main}} = \frac{1}{B_{\mathrm{E}} K} \sum_{b=1}^{B_{\mathrm{E}}} \sum_{k=1}^{K} \left(\widehat{y}_{b,k}^{\mathrm{main}} - y_{b,k} \right)^2 .$$

The HGA and average-fusion branches predict the same trajectory target through separate regression heads. Their losses, $\mathcal{L}_{\mathrm{HGA}}$ and $\mathcal{L}_{\mathrm{avg}}$, use the same mean-squared error as $\mathcal{L}_{\mathrm{main}}$ on predictions from F^{H} and the average-fusion representation, respectively. The complete ECoG adaptation objective is

$$\mathcal{L}_{\mathrm{ECoG}} = \mathcal{L}_{\mathrm{main}} + \lambda_{\mathrm{HGA}} \mathcal{L}_{\mathrm{HGA}} + \lambda_{\mathrm{avg}} \mathcal{L}_{\mathrm{avg}} + \lambda_{\mathrm{cf}} \mathcal{L}_{\mathrm{cf}}.$$

At inference, the three prediction branches are retained and combined using fixed convex weights. Appendix 11 provides the computational details and exact loss and prediction-combination weights.

4 Experiments

We systematically evaluate GeoLPE under cross-subject and within-subject EEG settings and further examine its EEG-to-ECoG transfer capability. Full fine-tuning and linear probing assess end-to-end adaptation and the linear decodability of frozen representations, respectively. Ablation studies examine the pretraining design, auxiliary objectives, and continuous geometric modeling, while representation analyses further characterize motor-imagery class structure across participants and between training and evaluation partitions.

4.1 Experimental Setup

Datasets and preprocessing. GeoLPE is pretrained from random initialization on 97 publicly available EEG dataset entries while preserving recording-specific electrode montages and three-dimensional coordinates. Downstream evaluation spans eight EEG datasets covering motor imagery, event-related potentials, c-VEP, and cognitive workload classification. For ECoG, we use Stanford FingerFlex recordings from nine participants with continuous five-finger flexion trajectories (Miller, 2019). EEG preprocessing follows REVE (El Ouahidi et al., 2025). Pretraining sources, downstream dataset statistics, and preprocessing details are provided in Appendix 7.1, Appendix Table 7, and Appendix 7, respectively.

Evaluation protocols. Cross-subject EEG evaluation follows a leave-one-subject-out (LOSO) protocol. Within-subject evaluation uses the official training and evaluation sets of BCIC-IV-2a for both full fine-tuning and linear probing. For the remaining seven datasets, each participant's examples are split into 30% for training and 70% for evaluation under both protocols. Full fine-tuning updates the encoder and classification head, whereas linear probing freezes the encoder and trains only the linear classification head. For the main EEG comparisons in Tables 1 and 2, we report the mean and standard deviation over five runs with different random seeds. For ECoG, training segments are pooled across participants, and performance is evaluated once on later held-out segments from the same participants.

Baselines and metrics. The EEG baselines are LaBraM, CBraMod, and REVE (Jiang et al., 2024; Wang et al., 2025; El Ouahidi et al., 2025), which we reevaluate from public checkpoints using identical data splits and adaptation protocols. Balanced accuracy (BAcc) serves as the primary EEG metric, with additional metrics reported in Appendix 12. For ECoG, we compare against published results for PLS, DeepFingerNet, HiLoFuseNet, and CORTEG (Sun et al., 2025; Tao et al., 2025; Yang et al., 2026). Performance is measured by Pearson correlation, averaged equally across the five fingers within each participant and then across the nine participants.

Table 1: Cross-subject LOSO results. Values are mean ± standard deviation of BAcc (%). Mean denotes the unweighted average across eight datasets. The encoder remains frozen throughout linear probing. Bold indicates the highest mean BAcc for each dataset and the cross-dataset average within each protocol.

Method	BCIC-IV-2a	BCIC-IV-2b	BNCI2015001	BNCI2014008	BNCI2014009	Castillos	Hinss2021	Kojima2024A	Mean
Full fine-tuning									
LaBraM	38.54 ± 0.99	74.31 ± 1.12	60.26 ± 0.71	68.94 ± 0.81	77.50 ± 0.89	86.94 ± 0.60	52.12 ± 1.37	63.45 ± 1.69	65.26
CBraMod	47.53 ± 0.75	72.01 ± 0.55	61.00 ± 0.29	68.36 ± 1.15	76.25 ± 0.53	83.75 ± 1.41	51.02 ± 0.92	63.72 ± 0.52	65.46
REVE	54.27 ± 1.20	72.41 ± 1.19	62.84 ± 0.93	68.84 ± 0.92	76.58 ± 0.73	85.56 ± 1.23	50.77 ± 0.59	69.04 ± 1.16	67.54
GeoLPE	**56.13 ± 0.96**	**76.38 ± 0.58**	**66.08 ± 0.62**	**69.35 ± 0.65**	**78.72 ± 0.73**	**88.19 ± 0.61**	**53.28 ± 1.21**	**69.44 ± 0.92**	**69.70**
Linear probing									
LaBraM	27.72 ± 0.25	59.17 ± 0.72	55.32 ± 0.89	53.36 ± 0.53	60.21 ± 0.56	39.17 ± 1.15	46.61 ± 1.63	54.10 ± 0.80	49.46
CBraMod	27.76 ± 0.71	57.02 ± 1.13	55.91 ± 0.83	55.31 ± 1.10	61.91 ± 1.13	48.06 ± 1.82	44.43 ± 1.15	55.53 ± 0.98	50.74
REVE	38.73 ± 0.62	67.39 ± 0.59	55.72 ± 0.84	61.96 ± 0.27	71.40 ± 0.69	78.19 ± 1.32	45.26 ± 0.86	63.36 ± 0.54	60.25
GeoLPE	**45.62 ± 0.26**	**67.88 ± 0.41**	**59.35 ± 0.80**	**63.68 ± 0.68**	**72.10 ± 0.65**	**80.97 ± 1.05**	**47.63 ± 0.52**	**63.41 ± 1.17**	**62.58**

Table 2: Within-subject EEG adaptation. Values are mean ± standard deviation of BAcc (%), with Mean averaged across eight datasets. Both full fine-tuning and linear probing use the official 1:1 train:evaluation split for BCIC-IV-2a and 30%/70% training/evaluation splits for the remaining datasets. Bold indicates the best result within each protocol.

Method	BCIC-2a	BCIC-2b	BNCI-15-1	BNCI-14-8	BNCI-14-9	Castillos	Hinss2021	Kojima2024A	Mean
Full fine-tuning									
LaBraM	37.54 ± 1.12	56.62 ± 0.76	56.58 ± 1.04	66.19 ± 0.72	71.86 ± 0.26	43.37 ± 1.11	45.63 ± 0.94	57.27 ± 0.56	54.38
CBraMod	40.56 ± 0.96	55.37 ± 1.00	57.97 ± 0.30	61.35 ± 0.67	61.46 ± 0.38	39.20 ± 0.80	49.34 ± 0.63	55.50 ± 0.92	52.59
REVE	45.88 ± 0.88	68.08 ± 0.81	**62.65 ± 0.26**	65.94 ± 0.71	73.42 ± 0.74	61.55 ± 0.69	47.97 ± 1.02	59.70 ± 1.21	60.65
GeoLPE	**48.54 ± 0.27**	**72.80 ± 0.34**	61.20 ± 0.10	**70.67 ± 0.57**	**77.25 ± 0.90**	**66.67 ± 0.86**	**49.43 ± 0.50**	**63.24 ± 0.98**	**63.73**
Linear probing									
LaBraM	32.56 ± 0.87	49.17 ± 0.36	51.50 ± 0.60	54.31 ± 0.88	56.34 ± 0.64	31.82 ± 0.45	37.67 ± 0.39	48.30 ± 1.07	45.21
CBraMod	31.55 ± 0.57	49.30 ± 1.04	48.85 ± 0.35	52.07 ± 0.25	55.17 ± 1.19	33.90 ± 1.50	43.82 ± 1.14	49.43 ± 0.69	45.51
REVE	32.96 ± 1.03	57.53 ± 0.28	48.09 ± 1.09	59.40 ± 0.41	64.90 ± 0.18	53.52 ± 0.56	41.02 ± 1.22	51.99 ± 0.30	51.18
GeoLPE	**38.10 ± 0.93**	**59.02 ± 0.91**	**54.08 ± 0.48**	**62.31 ± 0.80**	**70.50 ± 0.94**	**57.05 ± 0.92**	**48.53 ± 0.52**	**56.22 ± 1.01**	**55.73**

4.2 Cross-Subject and Within-Subject Evaluation

Cross-subject generalization. Table 1 shows that GeoLPE achieves the highest mean BAcc across all eight datasets under both full fine-tuning and linear probing, with cross-dataset averages of 69.70% and 62.58%, respectively. These gains span all four task families and remain evident with the encoder frozen.

Within-subject adaptation. Table 2 shows that GeoLPE achieves the highest mean BAcc on seven of the eight datasets under full fine-tuning and on all eight under linear probing, with cross-dataset averages of 63.73% and 55.73%, respectively. REVE performs better on BNCI2015001 under full fine-tuning, showing that the gains vary across datasets. Absolute performance is not directly comparable between the cross-subject and within-subject settings because they use different data splits. Figure 2 summarizes performance across the eight datasets under all four EEG evaluation settings.

4.3 EEG-to-ECoG Transfer

Table 3 shows that GeoLPE achieves a mean Pearson correlation of $r = 0.566 \pm 0.125$ on Stanford FingerFlex, numerically exceeding the published 0.554 ± 0.154 reported for pooled CORTEG. Because the baselines are not reevaluated under a unified implementation, the comparison is limited to published results. These results support the transfer of an EEG-pretrained encoder to ECoG finger-trajectory prediction on held-out segments from seen participants, rather than generalization to unseen participants.

Table 3: ECoG finger-trajectory prediction on Stanford FingerFlex. Values are mean Pearson $r \pm$ standard deviation across nine participants. CORTEG and GeoLPE use pooled training data from all participants and chronological test segments; baseline results follow the cited sources.

	PLS	DeepFingerNet	HiLoFuseNet	CORTEG	**GeoLPE**
Model	Linear	Task-specific	Dual-stream	backbone	GeoLPE encoder
Input representation	Multiband wavelets	Multiband wavelets	LFS + HGA	LFS + HGA	LFS + HGA
Training setup	Participant-specific	Participant-specific	Participant-specific	Pooled multi-participant	Pooled multi-participant
Mean Pearson r	0.402 ± 0.115	0.542 ± 0.129	0.534 ± 0.138	0.554 ± 0.154	$\mathbf{0.566 \pm 0.125}$

Table 4: Matched ablations across all eight downstream EEG datasets under LOSO evaluation. Values report macro BAcc (%), averaged across datasets. Full GeoLPE denotes the complete configuration; each remaining entry changes or removes the indicated component.

Ablation group	Variant I	Macro BAcc	Variant II	Macro BAcc	Variant III	Macro BAcc
Reference	Full GeoLPE	62.58	—	—	—	—
Pretraining design	Waveform target	59.94	Latent only	58.52	Random mask	61.08
Auxiliary objectives	w/o $\mathcal{L}_{\text{var}}$	58.29	w/o $\mathcal{L}_{\text{wave}}$	60.52	w/o $\mathcal{L}_{\text{boundary}}$	59.08
Geometry encoding	Time-only position	57.89	w/o Q/K rotation	58.58	w/o geometry bias	58.16

4.4 Ablation and Representation Analysis

Ablation study. Table 4 shows that the complete GeoLPE configuration achieves a macro BAcc of 62.58% under LOSO evaluation across the eight downstream datasets. Within the pretraining design group, replacing latent targets with waveform targets reduces performance to 59.94%, while removing the auxiliary objectives altogether yields 58.52%; random masking has a smaller effect at 61.08%. Removing any individual auxiliary objective also degrades performance, with the largest drop observed without $\mathcal{L}_{\text{var}}$ at 58.29%. Geometry ablations produce similarly consistent reductions: using time-only position gives 57.89%, while removing Q/K rotation or geometry bias yields 58.58% and 58.16%, respectively. Together, these results support the contributions of structured latent prediction, auxiliary objectives, and continuous geometric modeling.

Class structure across participants and data splits. We analyze frozen representations on BCIC-IV-2a using class-conditional subject-centroid dispersion D_{subj}, subject-conditional class-centroid separation D_{class}, and nearest-class-prototype readout. Lower D_{subj} indicates reduced subject-specific variation within classes, whereas higher D_{class} indicates stronger class separation within subjects.

Prototype readout evaluates transfer across participants under LOSO and from the training set to the evaluation set without gradient-based classifier training. In each setting, feature preprocessing and prototype estimation use only the reference participants or training data.
Table 5 shows that GeoLPE achieves the lowest mean D_{subj}, the highest mean D_{class}, and the highest prototype BAcc under both LOSO and training-to-evaluation protocols. Together, these results suggest reduced participant-specific dispersion and more transferable class structure. Figure 3 visualizes the class-conditional participant structure. Appendix Table 12 further shows that supervised adaptation on the training set increases evaluation-set prototype BAcc to 46.03%. Despite this improvement, the two-dimensional projections remain substantially overlapping, with silhouette scores near zero or negative.

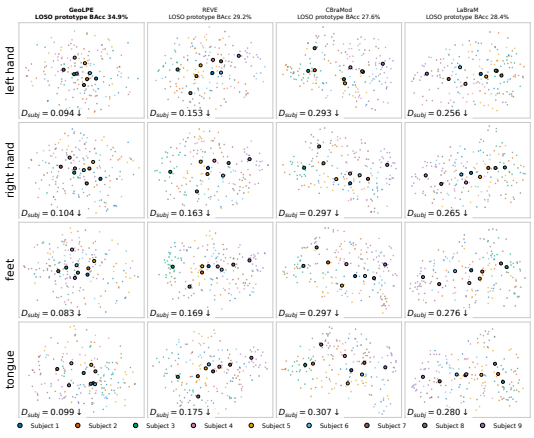

Figure 3: Class-conditional participant structure in BCIC-IV-2a representations. Colors denote participants and markers their centroids. Comparisons use D_{subj} in PCA(50) space.

Table 5: Centroid structure and prototype readout of frozen representations. D_{subj} and D_{class} are averaged across classes and participants, respectively; BAcc (%) is reported across participants. Prototypes are fitted on reference labels and evaluated on held-out participants or the test set.

Method	Mean $D_{\text{subj}}\downarrow$	LOSO prototype BAcc (%) ↑	Mean $D_{\text{class}}\uparrow$	prototype BAcc (%) ↑
GeoLPE	**0.095**	**34.93 ± 1.57**	**0.0199**	**40.35 ± 0.80**
REVE	0.165	29.21 ± 0.57	0.0104	30.56 ± 1.58
CBraMod	0.299	27.60 ± 1.51	0.0124	29.40 ± 0.84
LaBraM	0.269	28.41 ± 1.59	0.0113	29.36 ± 1.12

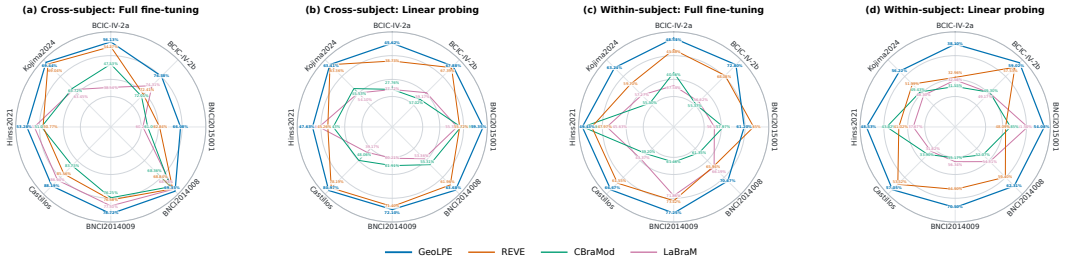

Figure 2: Visualization of EEG downstream performance across eight datasets. Panels (a) and (b) show cross-subject full fine-tuning and linear probing, respectively; panels (c) and (d) show the corresponding within-subject results. Labels report mean BAcc (%). Each dataset axis uses an independent local range, while the corresponding standard deviations are reported in Tables 1 and 2.

We therefore interpret these results as improved centroid-based readout across participants and the training and evaluation partitions, rather than fully separated class clusters. Metric definitions and participant-level analyses are provided in Appendix 13; Appendix Figures 4–6 show the corresponding training- and evaluation-set representation structure.

5 Conclusion

We presented GeoLPE, a geometry-aware latent-prediction framework for EEG foundation modeling. GeoLPE combines continuous absolute spatiotemporal encoding with explicit relative inter-electrode geometry, enabling a shared encoder to operate across heterogeneous electrode montages and recording lengths. Its structured joint-embedding objective places the primary pretraining supervision in latent space while retaining waveform reconstruction only as an auxiliary signal. Across eight EEG datasets, GeoLPE achieves the strongest average performance among the evaluated models under both cross-subject and within-subject evaluation, with consistent gains under full fine-tuning and linear probing. Ablations further support the contribution of continuous geometric modeling, while representation analyses show improved class-conditional alignment and prototype-based transfer across participants and from the training set to the evaluation set. Beyond EEG, the pretrained encoder can be reused for ECoG finger-trajectory prediction when combined with a complementary HGA pathway, demonstrating its applicability across recording modalities. Together, these results highlight continuous geometry and latent predictive pretraining as complementary ingredients for reusable neural signal representations.

6 Limitations

Our cross-modal transfer evaluation is limited to one ECoG dataset and motor-decoding task, using later segments from participants seen during training. It therefore does not establish transfer to unseen participants, electrode arrays, datasets, or sites. EEG-pretrained representations may also benefit stereoelectroencephalography (sEEG), which remains unevaluated. Future work will extend GeoLPE to multi-dataset ECoG and sEEG benchmarks, emphasizing unseen-participant and low-label transfer.

AI USE STATEMENT

In this work, we used generative AI tools for no tasks requiring mandatory disclosure. We have not used generative AI tools for any other tasks requiring mandatory disclosure, and the remaining mandatory-disclosure categories are not applicable to this work. Additionally, we used generative AI tools for language editing and stylistic refinement. We have reviewed all AI-assisted work. All AI-assisted edits were manually reviewed and revised by the authors, and generative AI was not used to generate scientific claims, methods, results, or citations. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work does not involve human-subject research or the collection of personal data. The authors have identified no specific ethical concerns associated with this study.

REPRODUCIBILITY STATEMENT

Appendix 7 summarizes the datasets and preprocessing procedures. Appendix 8 describes the evaluation protocols and training configurations. The remaining appendices provide algorithmic and computational details for reproducibility.

REFERENCES

P Aricò, F Aloise, F Schettini, S Salinari, D Mattia, and F Cincotti. Influence of P300 latency jitter on event related potential-based brain–computer interface performance. *Journal of Neural Engineering*, 11(3):035008, 2014. doi: 10.1088/1741-2560/11/3/035008.

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

Alexandre Barachant. Eeg Motor Imagery Dataset From The Phd Thesis "Commande Robuste D'Un Effecteur Par Une Interface Cerveau Machine Eeg Asynchrone", 2012.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mido Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. Featured Certification; continuous publication without an assigned volume, page range, or DOI.

Stephanie Brandl and Benjamin Blankertz. Motor Imagery Under Distraction— An Open Access BCI Dataset. *Frontiers in Neuroscience*, 14, 2020. doi: 10.3389/fnins.2020.566147.

Kalou Cabrera Castillos, Simon Ladouce, Ludovic Darmet, and Frédéric Dehais. Burst c-VEP Based BCI: Optimizing stimulus design for enhanced classification with minimal calibration data and improved user experience. *NeuroImage*, 284:120446, 2023. doi: 10.1016/j.neuroimage.2023.120446.

R. T. Canolty, E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628, 2006. doi: 10.1126/science.1128115.

Grégoire Cattan, Pedro L. C. Rodrigues, and Marco Congedo. EEG Alpha Waves dataset, 2018.

Grégoire Cattan, Anton Andreev, Pedro L. C. Rodrigues, and Marco Congedo. Dataset of an EEG-based BCI experiment in Virtual Reality and on a Personal Computer, 2019a.

Grégoire Cattan, Pedro L. C. Rodrigues, and Marco Congedo. Passive Head-Mounted Display Music-Listening EEG dataset, 2019b.

Juan David Chailloux Peguero, Omar Mendoza-Montoya, and Javier M. Antelis. Single-Option P300-BCI Performance Is Affected by Visual Stimulation Conditions. *Sensors*, 20(24):7198, 2020. doi: 10.3390/s20247198.

Wenwen Chang, Weixuan Kong, Guanghui Yan, Renjie Lv, Kaiyue Du, Muhammad Tariq Sadiq, Bin Guo, Rong Yin, and Xuan Liu. A multi-paradigm EEG dataset for studying upper limb rehabilitation exercises. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-06147-6.

Jingjing Chen, Dan Zhang, Andreas K. Engel, Qin Gong, and Alexander Maye. Application of a single-flicker online SSVEP BCI for spatial navigation. *PLOS ONE*, 12(5):e0178385, 2017. doi: 10.1371/journal.pone.0178385.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A Large Finer-grained Affective Computing EEG Dataset. *Scientific Data*, 10(1), 2023. doi: 10.1038/s41597-023-02650-w.

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

Markus R. Crell and Gernot R. Müller-Putz. Handwritten character classification from EEG through continuous kinematic decoding. *Computers in Biology and Medicine*, 182:109132, 2024. doi: 10.1016/j.compbiomed.2024.109132.

Yue Dong and Sen Tian. A large database towards user-friendly SSVEP-based BCI. *Brain Science Advances*, 9(4):297–309, 2023. doi: 10.26599/bsa.2023.9050020.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for eeg—adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems*, volume 38, pp. 22541–22577, 2025. doi: 10.52202/085713-0760.

Josef Faller, Carmen Vidaurre, Teodoro Solis-Escalante, Christa Neuper, and Reinhold Scherer. Autocalibration and Recurrent Adaptation: Towards a Plug and Play Online ERD-BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):313–319, 2012. doi: 10.1109/tnsre.2012.2189584.

Dylan Forenzo, Yixuan Liu, Jeehyun Kim, Yidan Ding, Taehyung Yoon, and Bin He. Integrating Simultaneous Motor Imagery and Spatial Attention for EEG-BCI Control. *IEEE Transactions on Biomedical Engineering*, 71(1):282–294, 2024. doi: 10.1109/tbme.2023.3298957.

Jing'ao Gao, Yao Liu, Zhengshuang Li, Kaixin Huang, Fan Wang, Jiaping Xu, Lei Zhao, Tianwen Li, and Yunfa Fu. An EEG Dataset for Visual Imagery-Based Brain–Computer Interface. *Scientific Data*, 13(1), 2026. doi: 10.1038/s41597-025-06512-5.

Alessandro T. Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M. Cichy. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264:119754, 2022. doi: 10.1016/j.neuroimage.2022.119754.

M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss. Beamforming in Noninvasive Brain–Computer Interfaces. *IEEE Transactions on Biomedical Engineering*, 56(4):1209–1219, 2009. doi: 10.1109/tbme.2008.2009768.

Eva Guttmann-Flury, Xinjun Sheng, and Xiangyang Zhu. Dataset combining EEG, eye-tracking, and high-speed video for ocular activity analysis across BCI paradigms. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-04861-9.

A. Hamid, K. Gagliano, S. Rahman, N. Tulin, V. Tchiong, I. Obeid, and J. Picone. The Temple University Artifact Corpus: An Annotated Corpus of EEG Artifacts. In *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–4. IEEE, 2020. doi: 10.1109/spmb50085.2020.9353647.

Yuheng Han, Yufeng Ke, Ruiyan Wang, Tao Wang, and Dong Ming. Enhancing SSVEP-BCI Performance Under Fatigue State Using Dynamic Stopping Strategy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:1407–1415, 2024. doi: 10.1109/tnsre.2024.3380635.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

Byeongho Heo, Song Park, Dongyoon Han, and Sangdoo Yun. Rotary position embedding for vision transformer. In *Computer Vision – ECCV 2024*, pp. 289–305, 2024. doi: 10.1007/978-3-031-72684-2_17.

Marcel F. Hinss, Emilie S. Jahanpour, Bertille Somon, Lou Pluchon, Frédéric Dehais, and Raphaëlle N. Roy. Open multi-session and multi-task EEG cognitive Dataset for passive brain-computer Interface Applications. *Scientific Data*, 10(1), 2023. doi: 10.1038/s41597-022-01898-y.

Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125, 2008. doi: 10.1016/j.jneumeth.2007.03.005.

Ping-Keng Jao, Ricardo Chavarriaga, Fabio Dell'Agnola, Adriana Arza, David Atienza, and Jose del R. Millan. EEG Correlates of Difficulty Levels in Dynamical Transitions of Simulated Flying and Mapping Tasks. *IEEE Transactions on Human-Machine Systems*, 51(2):99–108, 2021. doi: 10.1109/thms.2020.3038339.

Vinay Jayaram and Alexandre Barachant. MOABB: Trustworthy algorithm benchmarking for bcis. *Journal of Neural Engineering*, 15(6):066011, 2018. doi: 10.1088/1741-2552/aadea0.

Ji-Hoon Jeong, Jeong-Hyun Cho, Kyung-Hwan Shim, Byoung-Hee Kwon, Byeong-Hoo Lee, Do-Yeun Lee, Dae-Hyeok Lee, and Seong-Whan Lee. Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. *GigaScience*, 9 (10), 2020. doi: 10.1093/gigascience/giaa098.

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.

Emmanuel K. Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Eric Monacelli, and Yskandar Hamam. Online SSVEP-based BCI using Riemannian geometry. *Neurocomputing*, 191:55–68, 2016. doi: 10.1016/j.neucom.2016.01.007.

Emily S. Kappenman, Jaclyn L. Farrens, Wendy Zhang, Andrew X. Stewart, and Steven J. Luck. ERP CORE: An open resource for human event-related potential research. *NeuroImage*, 225: 117465, 2021. doi: 10.1016/j.neuroimage.2020.117465.

Murat Kaya, Mustafa Kemal Binli, Erkan Ozbay, Hilmi Yanar, and Yuriy Mishchenko. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific Data*, 5(1), 2018. doi: 10.1038/sdata.2018.211.

Heegyu Kim, Kyungho Won, Minkyu Ahn, and Sung Chan Jun. A 40-Class SSVEP Speller Dataset: Beta Range Stimulation for Low-Fatigue BCI Applications. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-06032-2.

Simon Kojima and Shin'ichiro Kanoh. An auditory brain-computer interface based on selective attention to multiple tone streams. *PLOS ONE*, 19(5):e0303565, 2024a. doi: 10.1371/journal.pone.0303565.

Simon Kojima and Shin'ichiro Kanoh. Four-class ASME BCI: investigation of the feasibility and comparison of two strategies for multiclassing. *Frontiers in Human Neuroscience*, 18, 2024b. doi: 10.3389/fnhum.2024.1461960.

Louis Korczowski, Martine Cederhout, Anton Andreev, Grégoire Cattan, Pedro Luis Coelho Rodrigues, Violette Gautheret, and Marco Congedo. Brain Invaders calibration-less P300-based BCI with modulation of flash duration Dataset (bi2015a), 2019a.

Louis Korczowski, Martine Cederhout, Anton Andreev, Grégoire Cattan, Pedro Luis Coelho Rodrigues, Violette Gautheret, and Marco Congedo. Brain Invaders Cooperative versus Competitive: Multi-User P300-based Brain-Computer Interface Dataset (bi2015b), 2019b.

Louis Korczowski, Ekaterina Ostaschenko, Anton Andreev, Grégoire Cattan, Pedro Luis Coelho Rodrigues, Violette Gautheret, and Marco Congedo. Brain Invaders calibration-less P300-based BCI using dry EEG electrodes Dataset (bi2014a), 2019c.

Louis Korczowski, Ekaterina Ostaschenko, Anton Andreev, Grégoire Cattan, Pedro Luis Coelho Rodrigues, Violette Gautheret, and Marco Congedo. Brain Invaders Solo versus Collaboration: Multi-User P300-based Brain-Computer Interface Dataset (bi2014b), 2019d.

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

Satyam Kumar, Hussein Alawieh, Frigyes Samuel Racz, Rawan Fakhreddine, and José del R Millán. Transfer learning promotes acquisition of individual BCI skills. *PNAS Nexus*, 3(2), 2024. doi: 10.1093/pnasnexus/pgae076.

Jongmin Lee, Minju Kim, Dojin Heo, Jongsu Kim, Min-Ki Kim, Taejun Lee, Jongwoo Park, HyunYoung Kim, Minho Hwang, Laehyun Kim, and Sung-Phil Kim. A comprehensive dataset for home appliance control using ERP-based BCIs with the application of inter-subject transfer learning. *Frontiers in Human Neuroscience*, 18, 2024. doi: 10.3389/fnhum.2024.1320457.

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), 2019. doi: 10.1093/gigascience/giz002.

Young-Eun Lee, Gi-Hwan Shin, Minji Lee, and Seong-Whan Lee. Mobile BCI dataset of scalp- and ear-EEGs with ERP and SSVEP paradigms while standing, walking, and running. *Scientific Data*, 8(1), 2021. doi: 10.1038/s41597-021-01094-4.

Robert Leeb, Felix Lee, Claudia Keinrath, Reinhold Scherer, Horst Bischof, and Gert Pfurtscheller. Brain–Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(4):473–482, 2007. doi: 10.1109/tnsre.2007.906956.

Bingchuan Liu, Xiaoshan Huang, Yijun Wang, Xiaogang Chen, and Xiaorong Gao. BETA: A Large Benchmark Database Toward SSVEP-BCI Application. *Frontiers in Neuroscience*, 14, 2020. doi: 10.3389/fnins.2020.00627.

Bingchuan Liu, Yijun Wang, Xiaorong Gao, and Xiaogang Chen. eldBETA: A Large Eldercare-oriented Benchmark Database of SSVEP-BCI for the Aging Population. *Scientific Data*, 9(1), 2022. doi: 10.1038/s41597-022-01372-9.

Haijie Liu, Penghu Wei, Haochong Wang, Xiaodong Lv, Wei Duan, Meijie Li, Yan Zhao, Qingmei Wang, Xinyuan Chen, Gaige Shi, Bo Han, and Junwei Hao. An EEG motor imagery dataset for brain computer interface in acute stroke patients. *Scientific Data*, 11(1), 2024a. doi: 10.1038/s41597-023-02787-8.

Xuan-Hao Liu, Yan-Kai Liu, Yansen Wang, Kan Ren, Hanwen Shi, Zilong Wang, Dongsheng Li, Bao-Liang Lu, and Wei-Long Zheng. EEG2Video: Towards Decoding Dynamic Visual Perception from EEG Signals. In *Advances in Neural Information Processing Systems 37*, NeurIPS 2024, pp. 72245–72273. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024b. doi: 10.52202/079017-2306.

Yuan Liu, Zhuolan Gui, De Yan, Zhuang Wang, Ruisi Gao, Ningxin Han, Junying Chen, Jialing Wu, and Dong Ming. Lower limb motor imagery EEG dataset based on the multi-paradigm and longitudinal-training of stroke patients. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-04618-4.

Xuelin Ma, Shuang Qiu, and Huiguang He. Multi-channel EEG recording during motor imagery of different joints from the same limb. *Scientific Data*, 7(1), 2020. doi: 10.1038/s41597-020-0535-2.

Víctor Martínez-Cagigal, Eduardo Santamaría-Vázquez, Sergio Pérez-Velasco, Diego Marcos-Martínez, Selene Moreno-Calderón, and Roberto Hornero. Non-binary m-sequences for more comfortable brain–computer interfaces based on c-VEPs. *Expert Systems with Applications*, 232: 120815, 2023. doi: 10.1016/j.eswa.2023.120815.

Kai J. Miller. A library of human electrocorticographic data and analyses. *Nature Human Behaviour*, 3(11):1225–1235, 2019. doi: 10.1038/s41562-019-0678-3.

Kai J. Miller, Marcel denNijs, Pradeep Shenoy, Jeffrey W. Miller, Rajesh P. N. Rao, and Jeffrey G. Ojemann. Spectral changes in cortical surface potentials during motor movement. *Journal of Neuroscience*, 27(9):2424–2432, 2007. doi: 10.1523/JNEUROSCI.3886-06.2007.

Andreas Miltiadous, Katerina D. Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G. Tsalikakis, Pantelis Angelidis, Markos G. Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, and Alexandros T. Tzallas. A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG. *Data*, 8(6):95, 2023. doi: 10.3390/data8060095.

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

Wajid Mumtaz, Syed Saad Azhar Ali, Mohd Azhar Mohd Yasin, and Aamir Saeed Malik. A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Medical & Biological Engineering & Computing*, 56(2):233–246, 2017. doi: 10.1007/s11517-017-1685-z.

Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials. *PLOS ONE*, 10(10):e0140703, 2015. doi: 10.1371/journal.pone.0140703.

Patrick Ofner, Andreas Schwarz, Joana Pereira, and Gernot R. Müller-Putz. Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PLOS ONE*, 12(8):e0182578, 2017. doi: 10.1371/journal.pone.0182578.

Patrick Ofner, Andreas Schwarz, Joana Pereira, Daniela Wyss, Renate Wildburger, and Gernot R. Müller-Putz. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. *Scientific Reports*, 9(1), 2019. doi: 10.1038/s41598-019-43594-9.

Steven M. Peterson, Zoe Steine-Hanson, Nathan Davis, Rajesh P. N. Rao, and Bingni W. Brunton. Generalized neural decoders for transfer learning across participants and recording modalities. *Journal of Neural Engineering*, 18(2):026014, 2021. doi: 10.1088/1741-2552/abda0b.

Hannah S Pulferer, Brynja Ásgeirsdóttir, Valeria Mondini, Andreea I Sburlea, and Gernot R Müller-Putz. Continuous 2D trajectory decoding from attempted movement: across-session performance in able-bodied and feasibility in a spinal cord injured participant. *Journal of Neural Engineering*, 19(3):036005, 2022. doi: 10.1088/1741-2552/ac689f.

Christoph Reichert, Igor Fabian Tellez Ceja, Catherine M. Sweeney-Reed, Hans-Jochen Heinze, Hermann Hinrichs, and Stefan Dürschmid. Impact of Stimulus Features on the Performance of a Gaze-Independent Brain-Computer Interface Based on Covert Spatial Attention Shifts. *Frontiers in Neuroscience*, 14, 2020. doi: 10.3389/fnins.2020.591777.

Angela Riccio, Luca Simione, Francesca Schettini, Alessia Pizzimenti, Maurizio Inghilleri, Marta Olivetti Belardinelli, Donatella Mattia, and Febo Cincotti. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Frontiers in Human Neuroscience*, 7, 2013. doi: 10.3389/fnhum.2013.00732.

Michele Romani, Devis Zanoni, Elisabetta Farella, and Luca Turchet. BrainForm: a Serious Game for BCI Training and Data Collection, 2025.

David Rozado, Andreas Duenser, and Ben Howell. Improving the Performance of an EEG-Based Motor Imagery Brain Computer Interface Using Task Evoked Changes in Pupil Diameter. *PLOS ONE*, 10(3):e0121262, 2015. doi: 10.1371/journal.pone.0121262.

G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004. doi: 10.1109/tbme.2004.827072.

Connor Schenck, Isaac Reid, Mithun George Jacob, Alex Bewley, Joshua Ainslie, David Rendleman, Deepali Jain, Mohit Sharma, Kumar Avinava Dubey, Ayzaan Wahid, Sumeet Singh, René Wagner, Tianli Ding, Chuyuan Fu, Arunkumar Byravan, Jake Varley, Alexey A. Gritsenko, Matthias Minderer, Dmitry Kalashnikov, Jonathan Tompson, Vikas Sindhwani, and Krzysztof Marcin Choromanski. Learning the RoPEs: Better 2D and 3D position encodings with STRING. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 53295–53315. PMLR, 2025.

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.

Andreas Schwarz, Carlos Escolano, Luis Montesano, and Gernot R. Müller-Putz. Analyzing and Decoding Natural Reach-and-Grasp Actions Using Gel, Water and Dry EEG Systems. *Frontiers in Neuroscience*, 14, 2020. doi: 10.3389/fnins.2020.00849.

Jian Shi, Danyang Chen, Xingwei Zhao, Zhixian Zhao, Shengjie Li, Yeguang Xu, Tao Ding, Zheng Zhu, Peng Zhang, Qing Ye, Yingxin Tang, Ping Zhang, Bo Tao, and Zhouping Tang. HEFMI-ICH: a hybrid EEG-fNIRS motor imagery dataset for brain-computer interface in intracerebral hemorrhage. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-06100-7.

Jaeyoung Shin, Alexander von Luhmann, Benjamin Blankertz, Do-Won Kim, Jichai Jeong, Han-Jeong Hwang, and Klaus-Robert Muller. Open Access Dataset for EEG+NIRS Single-Trial Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10): 1735–1745, 2017. doi: 10.1109/tnsre.2016.2628057.

Seyed Yahya Shirazi, Alexandre Franco, Maurício Scopel Hoffmann, Nathalia B. Esper, Dung Truong, Arnaud Delorme, Michael P. Milham, and Scott Makeig. HBN-EEG: The FAIR implementation of the Healthy Brain Network (HBN) electroencephalography dataset. *bioRxiv*, 2024. doi: 10.1101/2024.10.03.615261.

Marco Simões, Davide Borra, Eduardo Santamaría-Vázquez, Gbt-Upm, Mayra Bittencourt-Villalpando, Dominik Krzemiński, Aleksandar Miladinović, Neural_Engineering_Group, Thomas Schmid, Haifeng Zhao, et al. Bciaut-p300: A multi-session and multi-subject benchmark dataset on autism for p300-based brain-computer-interfaces. *Frontiers in Neuroscience*, 14:568104, 2020.

Arun Singh, Rachel C. Cole, Arturo I. Espinoza, Aron Evans, Scarlett Cao, James F. Cavanagh, and Nandakumar S. Narayanan. Timing variability and midfrontal ˜4 Hz rhythms correlate with cognition in Parkinson's disease. *npj Parkinson's Disease*, 7(1), 2021. doi: 10.1038/s41531-021-00158-x.

Jan Sosulski and Michael Tangermann. Electroencephalogram signals recorded from 13 healthy subjects during an auditory oddball paradigm under different stimulus onset asynchrony conditions, 2019.

Nitikorn Srisrisawang and Gernot R Müller-Putz. Simultaneous encoding of speed, distance, and direction in discrete reaching: an EEG study. *Journal of Neural Engineering*, 21(6):066042, 2024. doi: 10.1088/1741-2552/ada0ea.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Qiang Sun, Eva Calvo Merino, Bob Van Dyck, Yuan Yang, Jiayuan He, and Marc M. Van Hulle. Spectro-temporal fusion of high-gamma and low-frequency ECoG signals for intracranial finger movement decoding. *TechRxiv*, 2025. doi: 10.36227/techrxiv.176592065.57408366/v1.

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.

Wei Tao, Jianghao Hou, Yi Yang, Xun Chen, Tzyy-Ping Jung, and Feng Wan. DeepFingerNet predicts finger trajectory from ECoG measurements. *IEEE Transactions on Instrumentation and Measurement*, 74:1–12, 2025. doi: 10.1109/TIM.2025.3644562.

Mojgan Tavakolan, Zack Frehlick, Xinyi Yong, and Carlo Menon. Classifying three imaginary states of the same upper extremity using time-domain features. *PLOS ONE*, 12(3):e0174161, 2017. doi: 10.1371/journal.pone.0174161.

Nayid Triana-Guzman, Alvaro D. Orjuela-Cañon, Andres L. Jutinico, Omar Mendoza-Montoya, and Javier M. Antelis. Decoding EEG rhythms offline and online during motor imagery for standing and sitting based on a brain-computer interface. *Frontiers in Neuroinformatics*, 16, 2022. doi: 10.3389/fninf.2022.961089.

Erwan Vaineau, Alexandre Barachant, Anton Andreev, Pedro L. C. Rodrigues, Grégoire Cattan, and Marco Congedo. Brain Invaders Adaptive versus Non-Adaptive P300 Brain-Computer Interface dataset, 2018.

Gijsbrecht Franciscus Petrus Van Veen, Alexandre Barachant, Anton Andreev, Grégoire Cattan, Pedro Luis Coelho Rodrigues, and Marco Congedo. Building Brain Invaders: EEG data of an experimental validation, 2019.

Maitreyee Wairagkar, Yoshikatsu Hayashi, and Slawomir J. Nasuto. Exploration of neural correlates of movement intention based on characterisation of temporal dependencies in electroencephalography. *PLOS ONE*, 13(3):e0193722, 2018. doi: 10.1371/journal.pone.0193722.

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.

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.

Lu Wang, Zhenhao Zhang, Dan Han, Zhijun Zhang, Zhifang Liu, and Wei Liu. Single stimulus location for two inputs: A combined brain–computer interface based on Steady-State Visual Evoked Potential (SSVEP). *European Journal of Neuroscience*, 53(3):861–875, 2020. doi: 10.1111/ejn.15030.

Wei Wang, Fang He, Yifan Li, Wanying Qu, Yawei Li, Quanying Liu, and Yanwei Fu. Harnessing spectrum video for subject-level few-shot and cross-montage EEG generalization. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, 2026.

Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A Benchmark Dataset for SSVEP-Based Brain–Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1746–1752, 2017. doi: 10.1109/tnsre.2016.2627556.

Xiaoxi Wei, A. Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier, Vinay Jayaram, Camille Jeunet, Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas, Mehdi Bahri, Yannis Panagakis, Nikolaos Laskaris, Dimitrios A. Adamos, Stefanos Zafeiriou, William C. Duong, Stephen M. Gordon, Vernon J. Lawhern, Maciej Śliwowski, Vincent Rouanne, and Piotr Tempczyk. 2021 BEETL Competition: Advancing Transfer Learning for Subject Independence & Heterogenous EEG Data Sets, 2022.

Xiaoli Wu, Wenhui Zhang, Zhibo Fu, Roy T H Cheung, and Rosa H M Chan. An investigation of in-ear sensing for motor task classification. *Journal of Neural Engineering*, 17(6):066010, 2020. doi: 10.1088/1741-2552/abc1b6.

Ziqian Xie, Odelia Schwartz, and Abhishek Prasad. Decoding of finger trajectory from ECoG using deep learning. *Journal of Neural Engineering*, 15(3):036009, 2018. doi: 10.1088/1741-2552/aa9dbe.

Banghua Yang, Fenqi Rong, Yunlong Xie, Du Li, Jiayang Zhang, Fu Li, Guangming Shi, and Xiaorong Gao. A multi-day and high-quality EEG dataset for motor imagery brain-computer interface. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-04826-y.

Liuyin Yang, Qiang Sun, Bob Van Dyck, Eva Calvo Merino, and Marc M. Van Hulle. CORTEG: Foundation models enable cross-modality representation transfer from scalp to intracranial brain recordings. *arXiv preprint arXiv:2605.10337*, 2026. doi: 10.48550/arXiv.2605.10337.

Weibo Yi, Shuang Qiu, Kun Wang, Hongzhi Qi, Lixin Zhang, Peng Zhou, Feng He, and Dong Ming. Evaluation of EEG Oscillatory Patterns and Cognitive Process during Simple and Compound Limb Motor Imagery. *PLoS ONE*, 9(12):e114853, 2014. doi: 10.1371/journal.pone.0114853.

Weibo Yi, Jiaming Chen, Dan Wang, Xinkang Hu, Meng Xu, Fangda Li, Shuhan Wu, and Jin Qian. A multi-modal dataset of electroencephalography and functional near-infrared spectroscopy recordings for motor imagery of multi-types of joints from unilateral upper limb. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-05286-0.

Li Zheng, Sen Sun, Hongze Zhao, Weihua Pei, Hongda Chen, Xiaorong Gao, Lijian Zhang, and Yijun Wang. A Cross-Session Dataset for Collaborative Brain-Computer Interfaces Based on Rapid Serial Visual Presentation. *Frontiers in Neuroscience*, 14, 2020. doi: 10.3389/fnins.2020.579469.

Bangyan Zhou, Xiaopei Wu, Zhao Lv, Lei Zhang, and Xiaojin Guo. A Fully Automated Trial Selection Method for Optimization of Motor Imagery Based Brain-Computer Interface. *PLOS ONE*, 11(9):e0162657, 2016. doi: 10.1371/journal.pone.0162657.

Qing Zhou, Jiafan Lin, Lin Yao, Yueming Wang, Yan Han, and Kedi Xu. Relative Power Correlates With the Decoding Performance of Motor Imagery Both Across Time and Subjects. *Frontiers in Human Neuroscience*, 15, 2021. doi: 10.3389/fnhum.2021.701091.

Chongwen Zuo, Yi Yin, Haochong Wang, Zhiyang Zheng, Xiaoyan Ma, Yuan Yang, Jue Wang, Shan Wang, Zi-gang Huang, and Chaoqun Ye. Enhancing classification of a large lower-limb motor imagery EEG dataset for BCI in knee pain patients. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-05767-2.

Igor Zyma, Sergii Tukaev, Ivan Seleznov, Ken Kiyono, Anton Popov, Mariia Chernykh, and Oleksii Shpenkov. Electroencephalograms during Mental Arithmetic Task Performance. *Data*, 4(1):14, 2019. doi: 10.3390/data4010014.

7 Datasets and Preprocessing

7.1 Dataset Sources

Following the reporting format of prior large-scale EEG pretraining work (El Ouahidi et al., 2025), we provide the complete source list for the pretraining corpus used by the evaluated checkpoint. The frozen pretraining manifest contains 97 dataset entries: 88 entries loaded through MOABB (Jayaram & Barachant, 2018) and nine entries obtained from additional public repositories. An entry denotes one ingestion unit in the manifest rather than an independent cohort. Consequently, task-specific variants drawn from the same release are listed together and share the corresponding source citation.

Split	Modality	Entries	Role
Pretraining	EEG	88	MOABB entries
Pretraining	EEG	9	Additional public entries
Downstream	EEG	8	Classification datasets
Downstream	ECoG	1	Finger-trajectory regression

Table 6: Dataset provenance summary. Multiple entries can originate from a shared release and therefore do not necessarily represent independent cohorts. The complete source-to-citation mapping appears in Section 7.1.

MOABB entries (A–C). AlexMI (Barachant, 2012); Beetl2021_A and Beetl2021_B (Wei et al., 2022); BI2012 (Van Veen et al., 2019); BI2013a (Vaineau et al., 2018); BI2014a (Korczowski et al., 2019c); BI2014b (Korczowski et al., 2019d); BI2015a (Korczowski et al., 2019a); BI2015b (Korczowski et al., 2019b); BNCI2019_001 (Ofner et al., 2019); BNCI2020_001 (Schwarz et al., 2020); BNCI2020_002 (Reichert et al., 2020); BNCI2022_001 (Jao et al., 2021); BNCI2024_001 (Crell & Müller-Putz, 2024); BNCI2025_001 (Srisrisawang & Müller-Putz, 2024); BNCI2025_002 (Pulferer et al., 2022); Brandl2020 (Brandl & Blankertz, 2020); Cattan2019_PHMD (Cattan et al., 2019b); Cattan2019_VR (Cattan et al., 2019a); Chailloux2020 (Chailloux Peguero et al., 2020); Chang2025 (Chang et al., 2025); Chen2017SingleFlicker (Chen et al., 2017); and Cho2017 (Cho et al., 2017).

MOABB entries (D–L). Dong2023 (Dong & Tian, 2023); EPFLP300 (Hoffmann et al., 2008); ErpCore2021_ERN, ErpCore2021_LRP, ErpCore2021_MMN, ErpCore2021_N170, ErpCore2021_N2pc, ErpCore2021_N400, and ErpCore2021_P3 (Kappenman et al., 2021); Forenzo2023 (Forenzo et al., 2024); Gao2026 (Gao et al., 2026); GrosseWentrup2009 (Grosse-Wentrup et al., 2009); GuttmannFlury2025_ME, GuttmannFlury2025_MI, GuttmannFlury2025_P300, and GuttmannFlury2025_SSVEP (Guttmann-Flury et al., 2025); Han2024Fatigue (Han et al., 2024); HefmiIch2025 (Shi et al., 2025); Jeong2020 (Jeong et al., 2020); Kalunga2016 (Kalunga et al., 2016); Kaya2018 (Kaya et al., 2018); Kim2025BetaRange (Kim et al., 2025); Kojima2024B (Kojima & Kanoh, 2024b); Kumar2024 (Kumar et al., 2024); Lee2019_ERP, Lee2019_MI, and Lee2019_SSVEP (Lee et al., 2019); Lee2021Mobile_ERP and Lee2021Mobile_SSVEP (Lee et al., 2021); Lee2024_AC, Lee2024_BS, Lee2024_DL, Lee2024_EL, and Lee2024_TV (Lee et al., 2024); Liu2020BETA (Liu et al., 2020); Liu2022EldBETA (Liu et al., 2022); Liu2024 (Liu et al., 2024a); and Liu2025 (Liu et al., 2025).

MOABB entries (M–Z). Ma2020 (Ma et al., 2020); MartinezCagigal2023Pary (Martínez-Cagigal et al., 2023); Nakanishi2015 (Nakanishi et al., 2015); Ofner2017 (Ofner et al., 2017); PhysionetMI (Schalk et al., 2004); Rodrigues2017 (Cattan et al., 2018); RomaniBF2025ERP (Romani et al., 2025); Rozado2015 (Rozado et al., 2015); Schirrmeister2017 (Schirrmeister et al., 2017); Shin2017A and Shin2017B (Shin et al., 2017); Simoes2020 (Simões et al., 2020); Sosulski2019 (Sosulski & Tangermann, 2019); Tavakolan2017 (Tavakolan et al., 2017); TrianaGuzman2024 (Triana-Guzman et al., 2022); Wairagkar2018 (Wairagkar et al., 2018); Wang2016 (Wang et al., 2017); Wang2021Combined (Wang et al., 2020); Weibo2014 (Yi et al., 2014); Wu2020 (Wu et al., 2020); Yang2025 and Yang2025_3C (Yang et al., 2025); Yi2025 (Yi et al., 2025); Zheng2020 (Zheng et al., 2020); Zhou2016 (Zhou et al., 2016); Zhou2020 (Zhou et al., 2021); and Zuo2025 (Zuo et al., 2025).

Table 7: Downstream EEG datasets. Channels, durations, and sample counts reflect the inputs used in each task. For P300 datasets, each sample is a stimulus-locked ERP epoch. c-VEP denotes code-modulated visual evoked potentials.

Dataset	Paradigm	Channels	Duration	Samples	Classes	Participants
BCIC-IV-2a	Motor imagery	22	4.0 s	5,184	4	9
BCIC-IV-2b	Motor imagery	3	4.5 s	6,480	2	9
BNCI2015001	Motor imagery	13	5.0 s	5,600	2	12
BNCI2014008	ERP (P300)	8	1.0 s	33,600	2	8
BNCI2014009	ERP (P300)	16	0.8 s	17,280	2	10
Castillos	c-VEP	28	2.2 s	720	4	12
Hinss2021	Cognitive workload	61	2.0 s	13,380	3	15
Kojima2024A	Auditory ERP (P300)	64	1.7 s	11,200	2	11

Additional public entries. The remaining nine entries are NeuralBench_Chen2023Large (FACED) (Chen et al., 2023), NeuralBench_Gifford2022Large (THINGS-EEG2) (Gifford et al., 2022), NeuralBench_Hamid2020Tuar (TUAR) (Hamid et al., 2020), NeuralBench_Liu2024Eeg2video (SEED-DV) (Liu et al., 2024b), NeuralBench_Miltiadous2023Dice (Miltiadous et al., 2023), NeuralBench_Mumtaz2018Machine (Mumtaz et al., 2017), NeuralBench_Shirazi2024Hbn (HBN-EEG) (Shirazi et al., 2024), NeuralBench_Singh2021Timing (Singh et al., 2021), and NeuralBench_Zyma2019Electroencephalograms (EEGMAT) (Zyma et al., 2019).

Downstream EEG datasets. The eight downstream entries are BCIC-IV-2a (Tangermann et al., 2012), BCIC-IV-2b (Leeb et al., 2007), BNCI2015001 (Faller et al., 2012), BNCI2014008 (Riccio et al., 2013), BNCI2014009 (Aricò et al., 2014), Castillos (Cabrera Castillos et al., 2023), Hinss2021 (Hinss et al., 2023), and Kojima2024 (Kojima & Kanoh, 2024a).

ECoG dataset. We evaluate intracranial transfer on the Stanford FingerFlex recordings from the public ECoG library of Miller (Miller, 2019).

7.2 EEG Preprocessing

Each scalp EEG recording retains its observed electrode set and associated three-dimensional coordinates. Following REVE (El Ouahidi et al., 2025), signals undergo source-appropriate power-line suppression and band-pass filtering before being resampled to 200 Hz. Pretraining examples contain one to four seconds of continuous EEG, and longer recordings are randomly cropped. Within each crop, every channel is standardized using its median and median absolute deviation, after which amplitudes are clipped to $[-20, 20]$. The GeoLPE encoder partitions each channel into 100-sample waveform patches with an overlap of 10 samples.

Downstream evaluation retains the trial boundaries and labels supplied by each dataset. Sampling rate, input duration, and amplitude processing are converted to the common model input format. All compared encoders receive identical participant and trial partitions. Dataset-specific task windows and sample counts are reported in Table 7.

7.3 ECoG Preprocessing

For Stanford FingerFlex, unusable electrodes are removed before common-average referencing, and notch filters are applied at 60, 120, and 180 Hz. The low-frequency pathway uses signals band-pass filtered to 0.1–75 Hz and resampled to 200 Hz. The HGA pathway uses Hilbert amplitude envelopes from 70–110, 130–170, and 190–230 Hz. Each envelope is transformed by $\log(1+x)$ and standardized using channel statistics estimated only from the training interval; the three standardized envelopes are then averaged with equal weights.

Chronological split boundaries are fixed before window extraction, and no window crosses a split boundary. Each example contains one second of neural history and predicts five finger trajectories

at 25 Hz with a 40 ms neural-to-movement delay. Pearson correlation is averaged equally across the five fingers of each participant and then equally across participants.

8 Experimental Protocols and Implementation

8.1 EEG Evaluation Protocols

Cross-subject evaluation follows leave-one-subject-out (LOSO) validation. In each fold, one participant is reserved for testing, while the data from every remaining participant are divided by class into 80% training and 20% validation subsets. The held-out participant is not used for optimization or model selection. Full fine-tuning updates the encoder and classification head, whereas linear probing freezes the encoder and optimizes only the readout.

Within-subject evaluation uses the official training and evaluation sets for BCIC-IV-2a under both full fine-tuning and linear probing. For the remaining seven datasets, 30% of each participant's examples are used for training and 70% for evaluation under both protocols. The same participant and trial partitions are applied to GeoLPE and every EEG baseline. All GeoLPE results use one shared pretrained checkpoint rather than selecting a separate pretraining checkpoint for each downstream dataset.

8.2 ECoG Evaluation Protocol

Stanford FingerFlex is divided chronologically into training, validation, and test intervals. Training windows from all nine participants are pooled, model selection uses the chronologically held-out validation intervals, and final performance is evaluated once on later test intervals from the same participants. This protocol measures EEG-to-ECoG adaptation within the observed participant cohort and does not evaluate transfer to unseen ECoG participants.

The ECoG adaptation configuration is optimized for at most 100 epochs. We evaluate every two epochs and stop after 15 evaluations without a validation improvement of at least 10^{-4}. The final extension stage freezes the established ECoG network and updates only the HGA-conditioned modulation and the rank-4 residual adapters in the final two backbone blocks.

8.3 Implementation Summary

Table 8 collects the settings fixed for the reported experiments.

9 Continuous Geometry: Computational Details

This section expands the spatial attention operator in Eq. equation 1. Consider layer ℓ, attention head h, and one fixed temporal position. Let $q_i, k_i, v_i \in \mathbb{R}^{d_h}$ be the query, key, and value of electrode i, where d_h is even. We first center the electrode coordinates as $p_i^{\mathrm{c}} = p_i - C^{-1}\sum_{j=1}^{C} p_j$.

Learnable orthogonal feature basis. Each layer and head has an unconstrained matrix $\Gamma_{\ell,h} \in \mathbb{R}^{d_h \times d_h}$. Its skew-symmetric part and Cayley transform are

$$S_{\ell,h} = \tfrac{1}{2}(\Gamma_{\ell,h} - \Gamma_{\ell,h}^{\top}), \qquad O_{\ell,h} = (I + S_{\ell,h})^{-1}(I - S_{\ell,h}). \tag{2}$$

Because $S_{\ell,h}^{\top} = -S_{\ell,h}$, $O_{\ell,h}$ is orthogonal. The same basis is applied to the query and key, $q_i^{\circ} = O_{\ell,h} q_i$ and $k_i^{\circ} = O_{\ell,h} k_i$, while the value remains unchanged.

Continuous spatial rotations. We divide the d_h features into $d_h/2$ adjacent pairs. Pair m has a learnable spatial frequency $\omega_{\ell,h,m} \in \mathbb{R}^3$ and electrode-specific angle $\vartheta_{i,m} = (p_i^{\mathrm{c}})^{\top}\omega_{\ell,h,m}$. With $R(\vartheta) = \left[\begin{smallmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{smallmatrix}\right]$, the rotated pair is

$$q'_{i,(m)} = R(\vartheta_{i,m}) q^{\circ}_{i,(m)}, \qquad k'_{i,(m)} = R(\vartheta_{i,m}) k^{\circ}_{i,(m)}. \tag{3}$$

The frequencies jointly learn spatial directions and scales. Since the query and key use the same basis and frequency at every electrode, their relative rotation satisfies $R(\vartheta_{i,m})^{\top} R(\vartheta_{j,m}) = R((p_j -$

Table 8: Implementation and evaluation configuration.

Component	Configuration
Pretraining corpus	97 public EEG dataset entries.
EEG input	200 Hz; one- to four-second crops; 100-sample patches with 10-sample overlap; channel-wise median/MAD standardization and clipping to $[-20, 20]$; montage-homogeneous batches without padded electrodes.
GeoLPE encoder	Four factorized spatiotemporal Transformer layers; hidden dimension 128; four attention heads.
Masking	Temporal-band and spatial-electrode-cluster masks sampled with equal probability; exact 50% target masking ratio.
Latent predictor	Two Transformer layers with eight heads; six global heads and two local heads per layer; temporal radius two and four nearest spatial neighbors.
View perturbations	Context-only coordinate noise with standard deviation 0.0025 m; recording-level log-gain noise with standard deviation 0.15 and log-gain clipping to $[-0.5, 0.5]$.
JEPA objective	$\lambda_{\text{var}} = 0.05$, $\lambda_{\text{wave}} = 0.10$, and $\lambda_{\text{boundary}}^{\max} = 0.20$; boundary weight increases linearly from 0 to 0.20 over normalized progress $[0.35, 0.75]$.
Target update	Cosine EMA momentum from 0.996 to 1.0 over pretraining.
Pretraining optimization	AdamW; maximum batch size 64 with a 4,096-channel batch budget; learning rate 8×10^{-4}; weight decay 0.05; $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon_{\text{Adam}} = 10^{-8}$; BF16 autocast; gradient-norm clipping at 1.0; 25,000 updates; seed 42.
Hardware	One NVIDIA H100 GPU with 80 GB memory.
ECoG input	0.1–75 Hz low-frequency signal at 200 Hz; three HGA envelopes; one-second neural history; 25 Hz trajectory output; 40 ms neural-to-movement delay.
ECoG architecture	0.5-s low-frequency patches and 0.125-s HGA patches; fusion after backbone block 1; one four-head HGA temporal Transformer layer with feed-forward width $2D$ and zero dropout; rank-4 residual adapters with zero dropout in backbone blocks 3–4.
ECoG objective and output	$\lambda_{\text{HGA}} = 0.25$, $\lambda_{\text{avg}} = 0.25$, and $\lambda_{\text{cf}} = 0.10$ during base adaptation; fixed validation-selected output weights $0.375/0.375/0.25$ for the main, HGA, and average-fusion predictions.
ECoG optimization	AdamW; batch size 64; backbone learning rate 10^{-4}; new-module learning rate 10^{-3}; weight decay 0.01; BF16 autocast; five warm-up epochs; minimum learning rate 10^{-6}; at most 100 epochs; seed 42.
Software and provenance	Python 3.12.13; PyTorch 2.6.0 with CUDA 12.4; Transformers 4.56.2; EEG and ECoG implementation commits `287c5ca` and `7170c70`, respectively.

$p_i)^\top \omega_{\ell,h,m}$). The resulting query–key inner product therefore depends on the continuous relative displacement rather than the coordinate origin.

Geometric bias and attention. For query electrode i and key electrode j, define $\delta_{ij} = p_i - p_j$, $d_{ij} = \|\delta_{ij}\|_2$, and $d_{\max} = \max_{a,b} \|p_a - p_b\|_2$. With $\epsilon = 10^{-6}$, the pair feature $\psi_{ij}^{\text{geo}} = [\delta_{ij}/\max(d_{ij}, \epsilon); d_{ij}/\max(d_{\max}, \epsilon)] \in \mathbb{R}^4$ contains direction and normalized distance. A learned vector $w_{\ell,h} \in \mathbb{R}^4$ gives $b_{ij}^{\ell,h} = w_{\ell,h}^\top \psi_{ij}^{\text{geo}}$. The complete per-head computation is

$$
\begin{aligned}
a_{ij}^{\ell,h} &= \frac{(q_i')^\top k_j'}{\sqrt{d_h}} + b_{ij}^{\ell,h}, \\
\alpha_{ij}^{\ell,h} &= \frac{\exp(a_{ij}^{\ell,h})}{\sum_{j'=1}^{C} \exp(a_{ij'}^{\ell,h})}, \\
o_i^{\ell,h} &= \sum_{j=1}^{C} \alpha_{ij}^{\ell,h} v_j.
\end{aligned}
\tag{4}
$$

Thus, the spatial rotations modify query–key compatibility, the geometric bias modifies the attention logits, and neither operation transforms the values.

10 GeoLPE Pretraining: Computational Details

10.1 Views, Masks, and Information Flow

For record b, the target encoder receives the reference signal X_b and reference coordinates $\mathcal{P}_b$. Only the context view is perturbed. For electrode i, we sample $\xi_{b,i} \sim \mathcal{N}(0, \sigma_p^2 I_3)$ with $\sigma_p = 0.0025$ m and set $\widetilde{p}_{b,i} = p_{b,i} + \xi_{b,i}$. We also sample $\epsilon_b^{\text{gain}} \sim \mathcal{N}(0, \sigma_g^2)$ with $\sigma_g = 0.15$, set $\gamma_b = \exp(\text{clip}(\epsilon_b^{\text{gain}}, -0.5, 0.5))$, and use $\widetilde{X}_b = \gamma_b X_b$. The same gain is applied to every electrode and sample in a record.

Let C_b and T_b be the numbers of electrodes and temporal patches in record b, so $N_b = C_b T_b$ is its number of electrode–time positions. Each record masks exactly $\text{round}(0.5 N_b)$ positions while retaining nonempty target and visible sets. Temporal masking selects one contiguous interval and masks it across all electrodes. Spatial masking samples one electrode as a center, orders the remaining electrodes by Euclidean distance to that center, and masks the nearest electrodes across all temporal positions. If the target budget is not divisible by the relevant axis length, a randomly selected subset on the boundary completes the exact budget. The two mask types are sampled with equal probability.

The context encoder is evaluated with the target positions excluded from the key and value sets of both spatial and temporal attention. Consequently, a target waveform cannot influence any context representation. The target encoder processes the complete reference view under stop-gradient. For an index $n = (b, i, t)$, its normalized target is $z_n = \text{sg}(\text{LayerNorm}(E_{\bar{\theta}}(X_b, \mathcal{P}_b)_n))$. The predictor input equals the context representation at a visible position and $q_{\text{mask}} + e_{i,t}$ at a target position. The two-layer predictor has eight attention heads in each layer. Six heads are global. Two local heads use the union of a same-electrode temporal neighborhood of radius two patches and the four nearest electrodes at the same time position. The predictor returns $\widehat{z}_n$ on the full electrode–time lattice.

10.2 Pretraining Objectives

Let $\mathcal{I}_{\text{M}}$ and $\mathcal{I}_{\text{V}}$ denote all target and visible indices in a mini-batch, with sizes N_{M} and N_{V}. We write $\overline{z}_n = \text{LayerNorm}(\widehat{z}_n) \in \mathbb{R}^D$. The elementwise Smooth L1 penalty uses $\beta = 1$:

$$\rho_\beta(r) = \begin{cases} r^2/(2\beta), & |r| < \beta, \\ |r| - \beta/2, & |r| \geq \beta. \end{cases} \tag{5}$$

The main latent loss averages this penalty over target positions and feature dimensions:

$$\mathcal{L}_{\text{latent}} = \frac{1}{N_{\text{M}} D} \sum_{n \in \mathcal{I}_{\text{M}}} \sum_{d=1}^{D} \rho_1(\overline{z}_{n,d} - z_{n,d}). \tag{6}$$

Variance floor. For every feature dimension, we compute $\mu_d = N_{\text{M}}^{-1} \sum_{n \in \mathcal{I}_{\text{M}}} \overline{z}_{n,d}$ and $\sigma_d = [N_{\text{M}}^{-1} \sum_{n \in \mathcal{I}_{\text{M}}} (\overline{z}_{n,d} - \mu_d)^2]^{1/2}$. The variance regularizer is

$$\mathcal{L}_{\text{var}} = \frac{1}{D} \sum_{d=1}^{D} \max(0, 1 - \sigma_d). \tag{7}$$

It penalizes a predicted feature only when its batchwise standard deviation across target positions falls below one.

Query-based waveform recovery. Let $h_{b,u} \in \mathbb{R}^D$ be the final context representation at visible position $u \in \mathcal{V}_b$, and let $q_{\text{pool}} \in \mathbb{R}^D$ be the encoder's global pooling query. We summarize the visible context by $\pi_{b,u} = \text{softmax}_u(q_{\text{pool}}^\top h_{b,u} / \sqrt{D})$ and $\zeta_b = \sum_{u \in \mathcal{V}_b} \pi_{b,u} h_{b,u}$. For target index $n = (b, i, t)$, a training-only decoder predicts $\widehat{x}_n = f_{\text{wave}}(\zeta_b + e_n^{\text{ctx}})$, where e_n^{ctx} is the context-view positional embedding and f_{wave} is a two-layer $D \to 4D \to L$ mapping with an intermediate ReLU. The target x_n is the corresponding reference-view waveform patch, and

$$\mathcal{L}_{\text{wave}} = \frac{1}{N_{\text{M}} L} \sum_{n \in \mathcal{I}_{\text{M}}} \sum_{r=1}^{L} |\widehat{x}_{n,r} - x_{n,r}|. \tag{8}$$

Mask-boundary alignment. This objective aligns predictor and target representations at visible positions, emphasizing those closest to a target region. For a record with more than one electrode, let $\eta_b = \max(\mathrm{median}_i \min_{j\neq i} \|p_{b,i} - p_{b,j}\|_2, \epsilon_s)$ with $\epsilon_s = 10^{-4}$; for a single-electrode record, set $\eta_b = 1$. We define $\chi_{b,i,t} = [p_{b,i}/\eta_b; t] \in \mathbb{R}^4$. Let $\mathcal{I}_{\mathrm{M}}^{(b)}$ denote the target indices belonging to record b. For visible index n, let $d_n^{\mathrm{mask}} = \min_{n' \in \mathcal{I}_{\mathrm{M}}^{(b)}} \|\chi_n - \chi_{n'}\|_2$ and $\nu_n = \max(d_n^{\mathrm{mask}}, 1)^{-1/2}$. Then

$$\mathcal{L}_{\mathrm{boundary}} = \frac{1}{N_{\mathrm{V}}} \sum_{n \in \mathcal{I}_{\mathrm{V}}} \nu_n \left[\frac{1}{D} \sum_{d=1}^{D} \rho_1(\overline{z}_{n,d} - z_{n,d}) \right]. \tag{9}$$

This normalization makes one spatial nearest-neighbor step comparable to one temporal patch step when assigning boundary weights.

Weights and target update. With normalized progress $\tau \in [0, 1]$, the reported model uses $\lambda_{\mathrm{var}} = 0.05$, $\lambda_{\mathrm{wave}} = 0.10$, and

$$\lambda_{\mathrm{boundary}}(\tau) = 0.20 \ \mathrm{clip}\left(\frac{\tau - 0.35}{0.75 - 0.35}, 0, 1 \right). \tag{10}$$

After each optimizer update, the target parameters follow $\bar{\theta} \leftarrow \mu_{\mathrm{EMA}}(\tau)\bar{\theta} + [1 - \mu_{\mathrm{EMA}}(\tau)]\theta$, where

$$\mu_{\mathrm{EMA}}(\tau) = 1 - (1 - 0.996)\frac{1 + \cos(\pi\tau)}{2}. \tag{11}$$

Gradients update the context encoder, predictor, pooling query, and waveform decoder, but never the target encoder. The predictor and all auxiliary decoder components are discarded after pretraining.

Algorithm 1 GeoLPE pretraining.

Input: Training set $\mathcal{D}$, context encoder E_θ, target encoder $E_{\bar{\theta}}$, and predictor G_ψ
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ at progress τ **do**
    for each recording $b \in \mathcal{B}$ **do**
      Sample $\xi_{b,i} \sim \mathcal{N}(0, \sigma_p^2 I_3)$ and set $\widetilde{p}_{b,i} \leftarrow p_{b,i} + \xi_{b,i}$
      Sample $\epsilon_b^{\mathrm{gain}} \sim \mathcal{N}(0, \sigma_g^2)$
      $\gamma_b \leftarrow \exp\Big(\mathrm{clip}(\epsilon_b^{\mathrm{gain}}, -0.5, 0.5)\Big)$; $\widetilde{X}_b \leftarrow \gamma_b X_b$
      Sample $m_b \sim \mathrm{Uniform}\{\mathrm{temporal}, \mathrm{spatial}\}$
      $(\mathcal{M}_b, \mathcal{V}_b) \leftarrow \mathrm{Mask}(m_b, C_b, T_b)$ with $|\mathcal{M}_b| = \mathrm{round}(0.5 C_b T_b)$
      $H_b^{\mathrm{ctx}} \leftarrow E_\theta(\widetilde{X}_b, \widetilde{\mathcal{P}}_b; \mathcal{V}_b)$
      $Z_b^{\mathrm{tgt}} \leftarrow \mathrm{sg}(\mathrm{LayerNorm}(E_{\bar{\theta}}(X_b, \mathcal{P}_b)))$
      $u_n \leftarrow h_n^{\mathrm{ctx}}$ for $n \in \mathcal{V}_b$
      $u_n \leftarrow q_{\mathrm{mask}} + e_{i,t}$ for $n = (b, i, t) \in \mathcal{M}_b$
    end for
    $\widehat{Z} \leftarrow G_\psi(U)$ and $\overline{z}_n \leftarrow \mathrm{LayerNorm}(\widehat{z}_n)$
    Compute $\mathcal{L}_{\mathrm{latent}}$, $\mathcal{L}_{\mathrm{var}}$, $\mathcal{L}_{\mathrm{wave}}$, and $\mathcal{L}_{\mathrm{boundary}}$
    Combine the four objectives into $\mathcal{L}$
    Update all trainable parameters except $\bar{\theta}$ using $\nabla\mathcal{L}$
    $\bar{\theta} \leftarrow \mu_{\mathrm{EMA}}(\tau)\bar{\theta} + [1 - \mu_{\mathrm{EMA}}(\tau)]\theta$
  end for
  return E_θ

11 ECoG Adaptation: Computational Details

11.1 Dual-Rate Alignment and Fusion

At 200 Hz, the low-frequency pathway uses 100-sample patches and the HGA pathway uses 25-sample patches. Thus, the same one-second ECoG window produces $T_{\mathrm{H}} = 4T_{\mathrm{L}}$ temporal positions per electrode. Let $H^{\mathrm{L}} \in \mathbb{R}^{B_{\mathrm{E}} \times C \times T_{\mathrm{L}} \times D}$ be the output of the first pretrained block and $F^{\mathrm{H}} \in \mathbb{R}^{B_{\mathrm{E}} \times C \times T_{\mathrm{H}} \times D}$ the output of the HGA temporal Transformer. We align the low-frequency representation to the fine lattice by

$$F^{\mathrm{L}}_{b,i,4(t-1)+s} = H^{\mathrm{L}}_{b,i,t}, \qquad s \in \{1, 2, 3, 4\}. \tag{12}$$

This operation repeats representations rather than waveform samples and establishes a fixed four-to-one correspondence between the fine and coarse temporal lattices.

The coordinate residual uses both the common coordinate and its array-normalized form. Writing $\bar{p} = C^{-1}\sum_i p_i$ and $r_{\max} = \max(\max_i \|p_i - \bar{p}\|_2, 10^{-4})$, we define $\widehat{p}_i = (p_i - \bar{p})/r_{\max}$ and $c_i = f_{\mathrm{coord}}([p_i; \widehat{p}_i]) \in \mathbb{R}^D$. The mapping f_{coord} is a $6 \to D \to D$ multilayer perceptron with a GELU nonlinearity; its output layer is initialized to zero. The same c_i is added to all low-frequency and HGA patch embeddings from electrode i.

The aligned streams are fused through a gated HGA residual followed by HGA-conditioned feature modulation. The gate mapping f_{gate} applies LayerNorm to the concatenated $2D$-dimensional input and then a linear $2D \to D$ projection. Its weight is initialized to zero and its bias to -2. Using pointwise mappings over the final dimension,

$$
\begin{aligned}
g^{\mathrm{gate}} &= \mathrm{sigmoid}\big(f_{\mathrm{gate}}([F^{\mathrm{L}}; F^{\mathrm{H}}])\big)\,, \\
U &= F^{\mathrm{L}} + g^{\mathrm{gate}} \odot W_{\mathrm{H}}\,\mathrm{LayerNorm}(F^{\mathrm{H}}), \qquad (13) \\
[m^{\mathrm{scale}}; m^{\mathrm{shift}}] &= W_{\mathrm{mod}}\,\mathrm{LayerNorm}(F^{\mathrm{H}}), \\
F^{\mathrm{F}} &= U + \tanh(m^{\mathrm{scale}}) \odot \mathrm{LayerNorm}(F^{\mathrm{L}}) + m^{\mathrm{shift}}. \qquad (14)
\end{aligned}
$$

The HGA residual and modulation output projections, W_{H} and W_{mod}, are initialized to zero. Consequently, $F^{\mathrm{F}} = F^{\mathrm{L}}$ at the start of adaptation. The remaining pretrained blocks process F^{F} on the fine temporal lattice.

11.2 Cross-Frequency Latent Prediction

For each coarse low-frequency position, we average the four HGA representations on its matched fine temporal support,

$$
A^{\mathrm{H}}_{b,i,t} = \frac{1}{4}\sum_{s=1}^{4} F^{\mathrm{H}}_{b,i,4(t-1)+s}. \qquad (15)
$$

A pointwise predictor P_{cf}, implemented as LayerNorm followed by a $D \to D \to D$ multilayer perceptron with GELU, maps A^{H} into the low-frequency latent space. We write $\widehat{H}^{\mathrm{cf}} = P_{\mathrm{cf}}(A^{\mathrm{H}})$ and $Z^{\mathrm{L}} = \mathrm{sg}(\mathrm{LayerNorm}(H^{\mathrm{L}}))$. The training-only auxiliary objective is

$$
\mathcal{L}_{\mathrm{cf}} = \frac{1}{B_{\mathrm{E}} C T_{\mathrm{L}}} \sum_{b=1}^{B_{\mathrm{E}}} \sum_{i=1}^{C} \sum_{t=1}^{T_{\mathrm{L}}} \left(1 - \frac{(\widehat{H}^{\mathrm{cf}}_{b,i,t})^{\top} Z^{\mathrm{L}}_{b,i,t}}{\|\widehat{H}^{\mathrm{cf}}_{b,i,t}\|_2 \|Z^{\mathrm{L}}_{b,i,t}\|_2}\right). \qquad (16)
$$

The stopped low-frequency target prevents this objective from pulling the GeoLPE pathway toward HGA. The predictor is discarded at inference and does not constitute an additional trajectory head.

11.3 Residual Adaptation and Trajectory Objectives

The final two backbone blocks contain bottleneck residual adapters of rank four. Given the representation H in block ℓ, the adapter update is

$$
\begin{aligned}
&\mathcal{A}_\ell(H) = W^{\mathrm{up}}_\ell\,\mathrm{GELU}\big(W^{\mathrm{down}}_\ell\,\mathrm{LayerNorm}(H)\big)\,, \\
&W^{\mathrm{down}}_\ell \in \mathbb{R}^{4\times D}, \qquad W^{\mathrm{up}}_\ell \in \mathbb{R}^{D\times 4}.
\end{aligned} \qquad (17)
$$

W^{up}_ℓ is initialized to zero and the update is added after the corresponding pretrained block. In the final extension stage, the established ECoG parameters remain frozen while these adapters and the HGA-conditioned modulation in Eq. equation 14 are optimized.

All three trajectory heads use the same normalized target $y_b \in \mathbb{R}^K$, where $K = 5$. The HGA head aggregates F^{H} by concatenating mean pooling with learned attention pooling and outputs $\widehat{y}^{\mathrm{HGA}}_b$. The average-fusion path forms $F^{\mathrm{A}} = (F^{\mathrm{L}} + F^{\mathrm{H}})/2$, passes it through the same subsequent encoder blocks used by the main path, and uses an independent pooling and regression head to produce $\widehat{y}^{\mathrm{avg}}_b$.

Table 9: Additional metrics for binary EEG tasks. Each cell reports the mean ± standard deviation of AUROC / AU-PR over five runs with different random seeds, in percent.

Dataset	LaBraM	CBraMod	REVE	GeoLPE
LOSO, full fine-tuning				
BCIC-IV-2b	83.16 ± 0.72 / 82.66 ± 0.69	82.69 ± 0.42 / 82.51 ± 0.63	81.81 ± 0.43 / 81.68 ± 0.59	85.20 ± 0.94 / 85.32 ± 0.83
BNCI2015001	68.43 ± 1.46 / 68.06 ± 1.76	72.00 ± 1.40 / 70.24 ± 1.60	72.29 ± 1.41 / 71.34 ± 1.00	76.79 ± 1.11 / 75.52 ± 1.67
BNCI2014008	76.25 ± 1.22 / 75.07 ± 1.14	76.58 ± 1.48 / 75.73 ± 1.40	76.38 ± 1.14 / 75.11 ± 1.96	77.28 ± 1.67 / 76.08 ± 1.71
BNCI2014009	85.00 ± 1.81 / 84.55 ± 1.42	86.45 ± 1.41 / 86.28 ± 1.38	86.29 ± 1.72 / 86.25 ± 1.30	86.71 ± 1.07 / 86.64 ± 1.42
Kojima2024A	68.83 ± 0.20 / 52.83 ± 0.66	70.35 ± 0.49 / 54.48 ± 0.98	76.09 ± 0.30 / 63.61 ± 1.54	77.14 ± 0.63 / 64.98 ± 1.01
LOSO, linear probing				
BCIC-IV-2b	63.28 ± 0.94 / 63.57 ± 0.70	61.38 ± 0.87 / 60.18 ± 0.68	82.45 ± 0.77 / 82.10 ± 0.35	81.91 ± 0.34 / 81.66 ± 1.69
BNCI2015001	62.16 ± 1.24 / 61.17 ± 1.60	61.13 ± 0.72 / 61.46 ± 0.86	76.06 ± 1.71 / 75.59 ± 1.69	74.81 ± 1.14 / 73.79 ± 1.51
BNCI2014008	55.18 ± 1.79 / 54.78 ± 1.05	58.88 ± 1.09 / 57.30 ± 1.08	72.99 ± 1.96 / 71.79 ± 1.59	76.03 ± 1.47 / 75.31 ± 1.57
BNCI2014009	65.00 ± 1.47 / 64.80 ± 1.90	67.51 ± 1.11 / 67.83 ± 1.28	85.07 ± 1.92 / 84.72 ± 1.87	85.59 ± 1.79 / 85.40 ± 1.89
Kojima2024A	57.79 ± 1.52 / 40.36 ± 1.40	59.88 ± 0.46 / 43.55 ± 0.59	75.69 ± 0.98 / 60.04 ± 1.44	76.11 ± 0.26 / 61.98 ± 0.89
Within-subject, full fine-tuning				
BCIC-IV-2b	58.20 ± 1.21 / 58.06 ± 1.98	57.82 ± 1.30 / 58.07 ± 1.89	68.83 ± 1.66 / 67.86 ± 1.83	73.68 ± 1.64 / 73.25 ± 1.00
BNCI2015001	58.82 ± 1.40 / 59.06 ± 1.63	61.46 ± 1.10 / 61.70 ± 1.47	60.49 ± 1.84 / 61.42 ± 1.96	65.14 ± 1.64 / 65.21 ± 1.72
BNCI2014008	71.83 ± 1.55 / 71.89 ± 1.09	66.28 ± 1.76 / 67.14 ± 1.21	76.21 ± 0.72 / 76.76 ± 0.57	82.04 ± 1.20 / 82.30 ± 1.03
BNCI2014009	80.26 ± 1.78 / 79.61 ± 0.23	66.61 ± 0.39 / 64.96 ± 0.28	83.11 ± 1.03 / 83.28 ± 1.22	86.45 ± 1.45 / 86.61 ± 0.10
Kojima2024A	61.32 ± 1.04 / 45.75 ± 1.04	60.74 ± 1.12 / 43.85 ± 0.48	62.15 ± 0.09 / 45.70 ± 0.67	69.16 ± 0.30 / 55.21 ± 1.13
Within-subject, linear probing				
BCIC-IV-2b	55.46 ± 0.44 / 54.79 ± 0.20	57.77 ± 1.45 / 58.36 ± 1.17	74.54 ± 1.52 / 73.93 ± 1.90	69.42 ± 1.33 / 69.47 ± 1.31
BNCI2015001	59.29 ± 1.36 / 60.17 ± 1.19	61.63 ± 0.73 / 61.97 ± 1.23	68.19 ± 1.32 / 67.70 ± 1.12	63.68 ± 1.28 / 63.30 ± 1.21
BNCI2014008	55.78 ± 1.28 / 55.48 ± 1.14	53.60 ± 1.82 / 53.67 ± 1.52	71.21 ± 0.26 / 70.97 ± 0.67	78.63 ± 0.21 / 79.12 ± 0.30
BNCI2014009	58.81 ± 1.65 / 58.42 ± 1.52	57.82 ± 1.62 / 57.43 ± 1.02	76.94 ± 1.60 / 77.64 ± 1.15	83.06 ± 1.38 / 82.56 ± 1.00
Kojima2024A	56.68 ± 1.88 / 40.29 ± 1.82	57.43 ± 1.73 / 40.06 ± 1.94	68.55 ± 0.14 / 53.92 ± 1.20	67.79 ± 0.52 / 51.85 ± 1.13

Their losses are

$$\mathcal{L}_{\mathrm{HGA}} = \frac{1}{B_{\mathrm{E}} K} \sum_{b=1}^{B_{\mathrm{E}}} \sum_{k=1}^{K} (\widehat{y}_{b,k}^{\mathrm{HGA}} - y_{b,k})^2,$$
$$\mathcal{L}_{\mathrm{avg}} = \frac{1}{B_{\mathrm{E}} K} \sum_{b=1}^{B_{\mathrm{E}}} \sum_{k=1}^{K} (\widehat{y}_{b,k}^{\mathrm{avg}} - y_{b,k})^2. \tag{18}$$

Together with the main-path loss defined in Section 3, the base-adaptation objective is $\mathcal{L}_{\mathrm{ECoG}} = \mathcal{L}_{\mathrm{main}} + 0.25\mathcal{L}_{\mathrm{HGA}} + 0.25\mathcal{L}_{\mathrm{avg}} + 0.10\mathcal{L}_{\mathrm{cf}}$. At inference, the three predictions come from one checkpoint and one forward graph. A global convex combination is selected on the chronological validation split over a simplex grid with increments of $1/8$. The selected prediction is

$$\widehat{y}_b = 0.375\widehat{y}_b^{\mathrm{main}} + 0.375\widehat{y}_b^{\mathrm{HGA}} + 0.25\widehat{y}_b^{\mathrm{avg}}. \tag{19}$$

12 Additional EEG Evaluation Metrics

Balanced accuracy is the primary EEG metric in the main text. For binary datasets, we additionally report the area under the receiver-operating-characteristic curve (AUROC) and the area under the precision–recall curve (AU-PR). For multiclass datasets, we additionally report Cohen's κ and support-weighted F1 (F1w). Tables 9 and 10 retain the same data partitions and adaptation protocols as the corresponding BAcc results. Within each run, metrics are averaged over participants. All tabulated EEG metrics are reported in %, and the tables give the mean and standard deviation over five runs with different random seeds.

13 Frozen-Representation Centroid Analysis

13.1 Common Inputs and Feature Extraction

The analysis uses all 5,184 valid BCIC-IV-2a trials. GeoLPE, REVE, CBraMod, and LaBraM receive the same trials, the same 22 common electrodes, and inputs sampled at 200 Hz. Every encoder is placed in evaluation mode, all random augmentation and dropout are disabled, and all parameters remain frozen. For trial n and model m, the final encoder output is arranged as an electrode–time token grid $H_n^{(m)} \in \mathbb{R}^{C \times T_m \times D_m}$, where T_m and D_m denote the model-specific

Table 10: Additional metrics for multiclass EEG tasks. Each cell reports the mean ± standard deviation of Cohen's κ / weighted F1 over five runs with different random seeds; both metrics are multiplied by 100.

Dataset	LaBraM	CBraMod	REVE	GeoLPE
LOSO, full fine-tuning				
BCIC-IV-2a	18.05 ± 1.32 / 49.57 ± 0.17	30.04 ± 1.00 / 38.04 ± 1.26	39.03 ± 1.60 / 54.24 ± 1.78	41.51 ± 1.28 / 53.20 ± 0.09
CastillosCVEP	82.59 ± 0.80 / 86.80 ± 0.71	78.33 ± 1.88 / 83.66 ± 1.35	80.75 ± 1.64 / 83.08 ± 1.25	84.25 ± 0.81 / 88.09 ± 1.70
Hinss2021	28.18 ± 2.06 / 50.57 ± 0.39	26.53 ± 1.38 / 47.07 ± 1.89	26.15 ± 0.89 / 47.18 ± 0.92	29.92 ± 1.82 / 50.76 ± 1.53
LOSO, linear probing				
BCIC-IV-2a	3.63 ± 0.33 / 23.73 ± 0.03	3.68 ± 0.95 / 16.64 ± 0.07	18.31 ± 0.83 / 42.10 ± 0.68	27.49 ± 0.35 / 48.96 ± 1.37
CastillosCVEP	18.89 ± 1.53 / 36.59 ± 1.13	30.75 ± 2.43 / 45.31 ± 1.33	70.92 ± 1.76 / 84.86 ± 1.12	74.63 ± 1.40 / 85.77 ± 1.25
Hinss2021	19.91 ± 2.45 / 43.15 ± 1.65	16.64 ± 1.73 / 36.78 ± 1.81	17.89 ± 1.29 / 45.62 ± 1.42	21.44 ± 0.78 / 50.84 ± 0.06
Within-subject, full fine-tuning				
BCIC-IV-2a	16.72 ± 1.49 / 33.43 ± 0.37	20.75 ± 1.28 / 29.17 ± 1.43	27.84 ± 1.17 / 34.30 ± 1.62	31.39 ± 0.36 / 53.46 ± 0.72
CastillosCVEP	24.49 ± 1.48 / 36.18 ± 1.05	18.93 ± 1.07 / 29.13 ± 1.45	48.73 ± 0.92 / 59.47 ± 1.95	55.56 ± 1.15 / 63.53 ± 1.01
Hinss2021	18.44 ± 1.41 / 28.81 ± 0.76	24.01 ± 0.95 / 43.89 ± 1.61	21.95 ± 1.53 / 46.51 ± 0.82	24.14 ± 0.75 / 46.85 ± 0.96
Within-subject, linear probing				
BCIC-IV-2a	10.08 ± 1.16 / 29.05 ± 1.25	8.73 ± 0.76 / 25.69 ± 1.76	10.61 ± 1.37 / 37.76 ± 1.31	17.47 ± 1.24 / 40.02 ± 1.41
CastillosCVEP	9.09 ± 0.60 / 27.40 ± 1.58	11.87 ± 2.00 / 33.83 ± 1.06	38.03 ± 0.75 / 54.01 ± 1.54	42.73 ± 1.23 / 64.33 ± 1.90
Hinss2021	6.50 ± 0.59 / 41.98 ± 1.22	15.73 ± 1.71 / 42.51 ± 0.33	11.53 ± 1.83 / 50.92 ± 1.01	22.79 ± 0.78 / 41.69 ± 1.88

Class	GeoLPE	REVE	CBraMod	LaBraM
Left hand	**0.0945**	0.1530	0.2932	0.2563
Right hand	**0.1042**	0.1628	0.2972	0.2647
Feet	**0.0826**	0.1686	0.2972	0.2757
Tongue	**0.0994**	0.1748	0.3066	0.2797
Mean	**0.0952**	0.1648	0.2986	0.2691

Table 11: Class-conditional subject-centroid dispersion $D_{\mathrm{subj}}(k)$ on BCIC-IV-2a. Lower is better.

number of temporal tokens and hidden dimension. We first average over time within each electrode, $\bar{h}_{n,i}^{(m)} = T_m^{-1} \sum_{t=1}^{T_m} H_{n,i,t}^{(m)}$, producing C channel-aligned features. We then average these features over the electrodes to obtain the trial-level vector $x_n^{(m)} = C^{-1} \sum_{i=1}^{C} \bar{h}_{n,i}^{(m)} \in \mathbb{R}^{D_m}$. For descriptive geometry and visualization, each model is standardized separately over all trials and projected to 50 dimensions by PCA, yielding $z_n^{(m)}$. This all-trial PCA is not used to compute held-out BAcc. In every held-out evaluation, the standardizer and PCA are refitted using only the corresponding reference participants or training set.

13.2 Class-Conditional Subject-Centroid Dispersion

Let y_n and s_n denote the class and participant of trial n. For class k, let μ_k be the centroid over all trials in the class and $\mu_{s,k}$ the centroid for participant s within that class:

$$\mu_k = \frac{1}{N_k} \sum_{n:y_n=k} z_n, \qquad \mu_{s,k} = \frac{1}{N_{s,k}} \sum_{n:y_n=k,\, s_n=s} z_n. \tag{20}$$

The normalized subject-centroid dispersion is

$$D_{\mathrm{subj}}(k) = \frac{\frac{1}{S} \sum_{s=1}^{S} \|\mu_{s,k} - \mu_k\|_2^2}{\frac{1}{N_k} \sum_{n:y_n=k} \|z_n - \mu_k\|_2^2}, \qquad S = 9. \tag{21}$$

The denominator is the total trial-level scatter within class k. The ratio therefore measures the fraction associated with participant-centroid shifts rather than the absolute scale of an encoder's features. Lower values indicate weaker participant-specific displacement conditional on task class. Table 11 reports every class separately.

GeoLPE has the lowest dispersion for left-hand, right-hand, feet, and tongue imagery. Averaged equally across the four classes, its value is 0.0952, a 42.2% reduction relative to REVE (0.1648), the lowest-dispersion baseline. Figure 3 visualizes the same conditioning. For each model, we

randomly select 20 trials for every participant–class pair, yielding a balanced subset of 720 trials, and apply t-SNE to their 50-dimensional PCA representations. Each row fixes the task class, point color identifies the participant, and the enlarged point is that participant's class centroid. The resulting two-dimensional layout is qualitative; all reported dispersion values use the retained 50-dimensional PCA representation and all 5,184 trials.

13.3 Participant-Conditional Class-Centroid Separation

For participant s, let μ_s denote the centroid over all of that participant's trials and $\mu_{s,k}$ the centroid of class k. We define

$$D_{\text{class}}(s) = \frac{\frac{1}{K}\sum_{k=1}^{K}\|\mu_{s,k} - \mu_s\|_2^2}{\frac{1}{N_s}\sum_{n:s_n=s}\|z_n - \mu_s\|_2^2}, \qquad K = 4. \tag{22}$$

This ratio measures the fraction of within-participant variation associated with the separation of class centroids. GeoLPE has a participant-averaged D_{class} of 0.0199, compared with 0.0104 for REVE, 0.0124 for CBraMod, and 0.0113 for LaBraM. It is higher than REVE and LaBraM for all nine participants and higher than CBraMod for eight of nine participants.

13.4 Held-Out Nearest-Prototype Evaluation

The LOSO analysis fits the standardizer and PCA on eight reference participants in each fold. Their task labels estimate one prototype per class, and each trial from the ninth participant is assigned to the nearest prototype in Euclidean distance. The training–evaluation analysis is performed separately for each participant: the standardizer, PCA, and four class prototypes are fitted only on the training set and then evaluated on the evaluation set. Both protocols freeze the encoder and do not optimize a parameterized classifier. They remain supervised nonparametric evaluations because training labels determine the class prototypes. Participant-level BAcc values, rather than individual trials, are the units used for aggregation.

13.5 Held-Out Evaluation-Set Geometry Before and After Training-Set Adaptation

Figures 4–6 use two compact panel annotations. The symbol s denotes the four-class silhouette coefficient computed in the retained PCA space for the model, participant, encoder state, and data split shown in that panel. Larger values indicate stronger class separation, while values near or below zero indicate substantial overlap. The displayed two-dimensional coordinates are only the first two principal components and are not used to compute s. In evaluation-set panels, `proto` denotes balanced accuracy (%) obtained by fitting four nearest-class prototypes on the training set and applying them to the evaluation set. We omit `proto` from the training-set reference because evaluating the prototypes on the samples used to estimate them would be a resubstitution score.

For each model and participant, Figures 4 and 5 share the same frozen training-set-fitted standardizer, PCA basis, and panel limits. Figure 6 uses a separate training-set-fitted transform after supervised adaptation and displays only the held-out evaluation-set representations. Because the frozen and adapted latent spaces use separately fitted transforms, absolute point locations should not be compared across encoder states.

Table 12 reports the high-dimensional quantities shown in the two held-out evaluation-set figures. Values are participant means $\pm$ sample standard deviations. The silhouette describes latent geometry after applying the corresponding training-set-fitted transform; the prototype BAcc is the held-out prediction metric.

With frozen encoders, GeoLPE attains a training-to-evaluation prototype BAcc of 40.35%, compared with 30.56% for REVE, 29.40% for CBraMod, and 29.36% for LaBraM. After training-set adaptation, the corresponding values increase to 46.03%, 37.31%, 36.81%, and 35.46%. GeoLPE exceeds REVE, CBraMod, and LaBraM for 6/9, 7/9, and 9/9 participants, respectively. Its adapted evaluation-set silhouette is also closest to zero, with $s = -0.003$ versus -0.029, -0.055, and -0.043 for the baselines, but remains near zero. We therefore interpret the adapted representation as making class information more consistently recoverable on held-out evaluation data, not as producing completely separated clusters.

Table 12: Class geometry on the official BCIC-IV-2a evaluation set before and after participant-specific adaptation using the official training set. The classification head is excluded from every representation. s_{eval} is the four-class silhouette in the retained PCA space. Prototype BAcc (%) fits class centroids on the official training set and evaluates them on the official evaluation set. Values are reported as mean $\pm$ standard deviation over five runs with different random seeds.

Encoder state	Model	s_{eval}	Training→evaluation prototype BAcc
Frozen	**GeoLPE**	-0.0068 ± 0.0022	40.35 ± 0.80
Frozen	REVE	-0.0116 ± 0.0037	30.56 ± 1.58
Frozen	CBraMod	-0.0185 ± 0.0032	29.40 ± 0.84
Frozen	LaBraM	-0.0148 ± 0.0028	29.36 ± 1.12
Adapted	**GeoLPE**	-0.0025 ± 0.0073	46.03 ± 1.01
Adapted	REVE	-0.0286 ± 0.0313	37.31 ± 1.11
Adapted	CBraMod	-0.0551 ± 0.0524	36.81 ± 0.82
Adapted	LaBraM	-0.0434 ± 0.0226	35.46 ± 0.43

R2 and R3 provide the frozen training-set and held-out evaluation-set reference, respectively, while R4 evaluates the evaluation set after supervised training-set adaptation. GeoLPE retains the highest held-out prototype BAcc both before and after adaptation. Because the frozen and adapted panels use separately fitted latent-space transforms, the evidence concerns transfer of a simple training-set class-centroid rule rather than invariance of raw two-dimensional coordinates across encoder states.

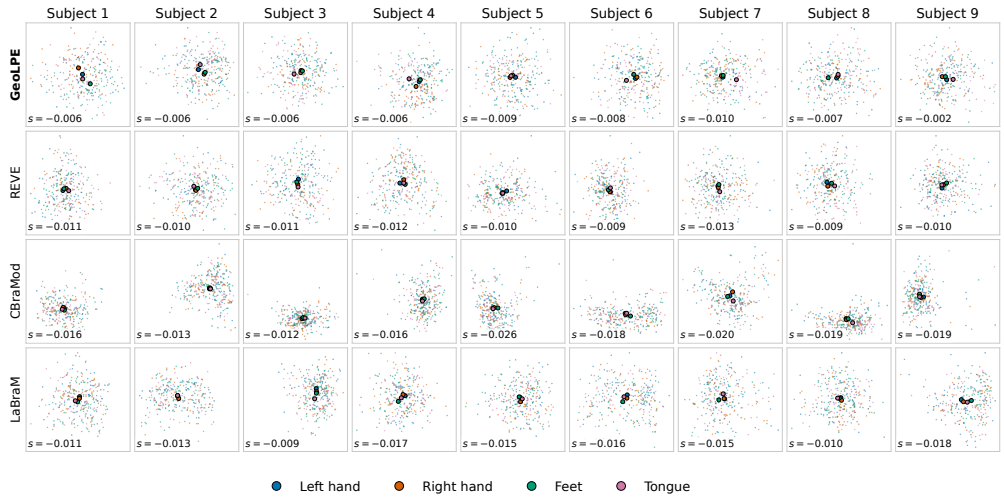

Figure 4: Frozen class geometry on the official BCIC-IV-2a training set. Rows identify encoders, columns identify participants, colors denote the four motor-imagery classes, and enlarged markers denote class centroids. Each panel reports the high-dimensional class silhouette s.

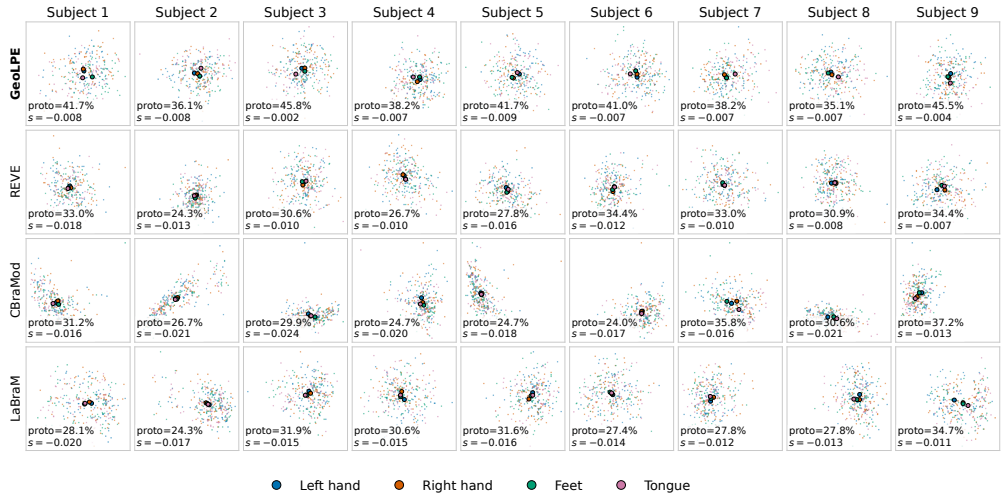

Figure 5: Frozen class geometry on the official BCIC-IV-2a evaluation set. The standardizer, PCA basis, and class prototypes are fitted only on the official training set; the evaluation set is transformed and evaluated out of sample. The annotations report training-to-evaluation nearest-prototype BAcc denoted by `proto` and high-dimensional class silhouette s.

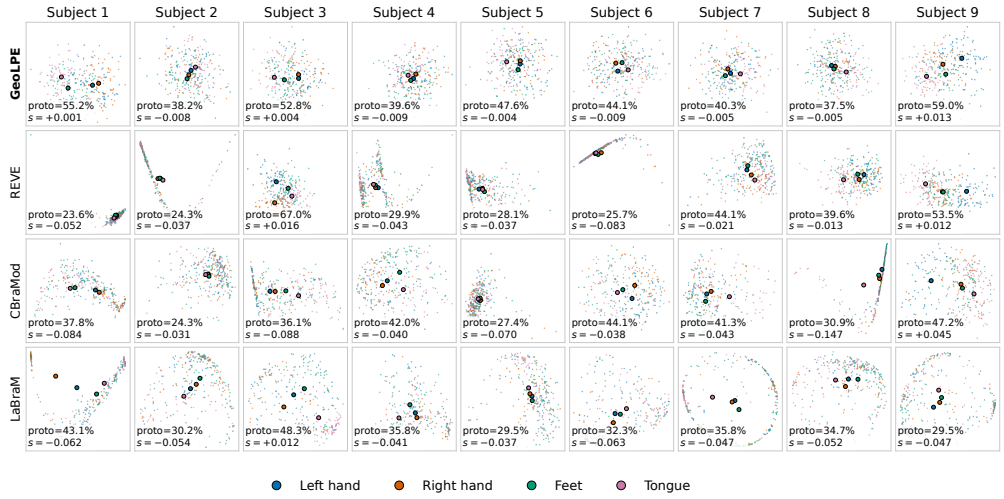

Figure 6: Evaluation-set class geometry after participant-specific adaptation on the official BCIC-IV-2a training set. The adapted encoder, standardizer, PCA basis, and class prototypes are fitted without evaluation-set data. Evaluation-set labels are used only for coloring and post-hoc evaluation. Annotations report training-to-evaluation nearest-prototype BAcc denoted by `proto` and high-dimensional class silhouette s.