HYPERBOLIC RELATION-AWARE GRAPH POOLING FOR EEG FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalography (EEG) is widely used in brain-computer interfaces, and large-scale pretraining has made EEG foundation models increasingly useful for downstream tasks. However, common readout heads pool pretrained tokens without explicitly considering which tokens convey similar information or which encode more specific semantic content. We propose Hyperbolic Relation-Aware Graph Pooling (HRAGP), a lightweight plug-and-play readout module that models these relations for downstream classification. HRAGP first maps EEG tokens into hyperbolic space to capture their geometric structure. Then, it constructs a weighted directed graph. Wherein, geodesic distance provides similarity-based edge weights, while entailment-cone compatibility, directional preference, and radial ordering determine which edges are retained and how they are oriented. Finally, a directional graph summarizes incoming and outgoing relations and combines the resulting representation with the average-pooled backbone features. Experiments with four EEG foundation models across five downstream benchmarks show that HRAGP improves over average pooling in most settings while adding relatively few trainable parameters.

1 INTRODUCTION

Electroencephalography (EEG) provides a non-invasive means of recording brain electrical activity with high temporal resolution, and has been widely used in Brain–Computer Interfaces (BCIs). However EEG decoding remains challenging due to its low signal-to-noise ratio, non-stationarity and variability across subjects and recording conditions. Hence traditional EEG decoders have been developed as task specific models. EEGNet (Lawhern et al., 2018), e.g., introduced a lightweight convolutional architecture applicable to multiple BCI paradigms, while EEG Conformer (Song et al., 2023) combined convolutional feature extraction with self-attention to capture longer-range dependencies. BENDR (Kostas et al., 2021) subsequently explored large-scale self-supervised pre-training for transferable EEG representations. Meanwhile geometric EEG learning has moved beyond Euclidean spaces through SPD-manifold attention (Pan et al., 2022), correlation-manifold modeling (Hu et al., 2025), and expanded to hyperbolic modeling (Li et al., 2026). These developments have shifted EEG decoding to more general representation learning.

Large-scale pre-training has accelerated this transition. Recent EEG foundation models (EEG-FMs) learn reusable representations from large collections of recordings. BIOT (Yang et al., 2023) unified heterogeneous biosignals through a shared Transformer encoder; LaBraM (Jiang et al., 2024) introduced large-scale masked pre-training for generic EEG representations; and EEGPT (Wang et al., 2024) used self-supervised spatio-temporal representation alignment for transferable EEG encoding. More recent models, including CBraMod (Wang et al., 2025) and REVE (El Ouahidi et al., 2025), further improve spatial-temporal modeling and adaptation across heterogeneous acquisition settings. Together with systematic evaluation efforts such as EEG-FM-Bench (Xiong et al., 2025), these models provide strong pretrained backbones for downstream EEG tasks.

By comparison, the downstream classifier has received considerably less attention. Most EEG-FMs still use a class token, average pooling, or a shallow MLP head before classification. EEG-FM-Bench (Xiong et al., 2025) shows that the choice of classifier configuration can materially affect downstream performance, suggesting that stronger representations do not determine the quality of adaptation. RHOP (Hu et al., 2026) makes this issue explicit by replacing the conventional classifi-

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.

cation head with Riemannian high-order pooling. Results show that pretrained EEG tokens contain useful information that may not be fully captured by standard readouts. High-order pooling mainly summarizes the global statistics of a token set, but it does not explicitly describe the sample-specific relations among tokens. This leaves room for a downstream head that models the internal relational structure of pretrained EEG representations.

Following the standard Gromov hyperbolicity measure (Gromov, 1987), we evaluate the normalized hyperbolicity of token representations across backbones and datasets. As shown in Fig 1, the token spaces exhibit a measurable degree of tree-like structure, motivating the use of hyperbolic geometry. We propose **Hyperbolic Relation-Aware Graph Pooling (HRAGP)**, a relation-aware downstream head for EEG-FMs, which combines geometric quantities to infer weighted, directional relations among tokens and summarizes the resulting relational structure for downstream prediction.

Our contributions are summarized as follows:

- **Hyperbolic EEG-FMs adaptation.** We introduce hyperbolic relation modeling into the downstream fine-tuning of EEG-FMs and provide evidence that pretrained EEG token spaces exhibit measurable tree-like structure across multiple backbones and datasets, motivating hyperbolic modeling of token relations.
- **Relation-aware graph pooling.** We propose HRAGP, which maps pretrained EEG-FMs tokens into hyperbolic space, induces a sample-specific weighted directed graph using geodesic proximity, entailment compatibility, and radial ordering. This is followed by the summarization of the resulting structure through directional graph readout.
- **Evaluation on multi-tasks.** We evaluate HRAGP on four EEG-FMs and five downstream benchmarks, achieving consistent improvements over average pooling. Besides, several ablation studies further isolate the effects of hyperbolic relation modeling, geometric relation conditions, and directional graph modeling.

2 Related Work

EEG-FMs and Downstream Adaptation EEG-FMs have improved transferable representation learning through large scale pre-training, including BENDR (Kostas et al., 2021), BIOT (Yang et al., 2023), LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024), CBraMod (Wang et al., 2025), and REVE (El Ouahidi et al., 2025). Benchmarks and surveys also provide unified evaluations of these models across datasets, adaptation protocols, and downstream tasks (Xiong et al., 2025; Liu et al., 2026; Kuruppu et al., 2026). These studies mainly focus on the pretrained encoder and its transfer ability, while the downstream readout is usually kept simple. EEG-FM-Bench (Xiong et al., 2025) shows that downstream performance can vary with the adaptation, and RHOP (Hu et al., 2026) further improves EEG-FMs transfer by exploiting high-order token statistics. HRAGP follows the direction, but focuses on sample-specific token relations rather than global distributional statistics.

Geometry-Aware and Hyperbolic Relation Modeling Riemannian methods have been widely used to preserve non-Euclidean structure. SPDNet (Huang & Van Gool, 2017) and its extensions (Brooks et al., 2019) established operations on SPD manifolds, while following methods such as MAtt (Pan et al., 2022) and CorAtt (Hu et al., 2025) exploit manifold geometry for correlation-based EEG models. Hyperbolic geometry has also been explored for EEG in HEEGNet (Li et al., 2026). Hyperbolic models provide compact representations for hierarchical data (Nickel & Kiela, 2017; 2018), while hyperbolic neural networks extend feature transformation and message passing to non-Euclidean spaces (Ganea et al., 2018b; Chami et al., 2019; Liu et al., 2019). To improve relation representations, hyperbolic entailment cones (Ganea et al., 2018a; Hasan et al., 2026) are used to encode asymmetric compatibility through directed geometric regions.

3 Preliminaries

We briefly review the hyperbolic geometry used in HRAGP. We adopt the Lorentz model for its convenient closed-form operations and stable optimization (Nickel & Kiela, 2018).

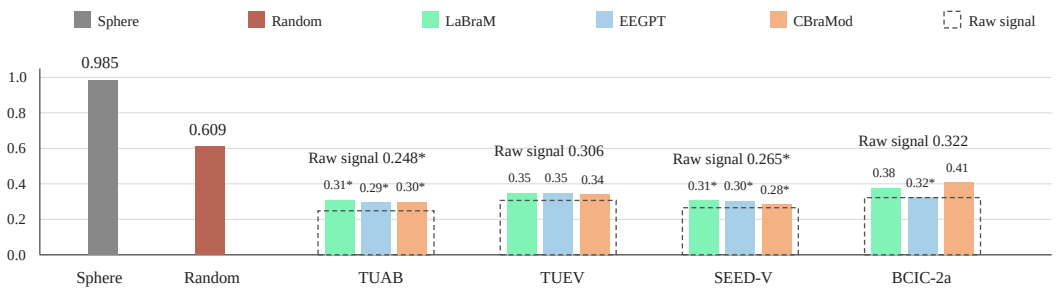


Figure 1: Normalized four-point Gromov hyperbolicity ($\delta_{\max}/D$) of raw EEG signals and token representations across four datasets. Dashed lines indicate the raw-signal value for each dataset; colored bars show the values for three EEG foundation models, with error bars denoting sample standard deviations. Sphere and Random provide reference values. Asterisks mark lower-bound estimates obtained from sampled quadruples. Lower values indicate more tree-like geometry.

Lorentz Model of Hyperbolic Space For $x, y \in \mathbb{R}^{d+1}$, the Lorentzian inner product is given by $\langle x, y\rangle_{\mathcal{L}} = -x_0 y_0 + \sum_{k=1}^{d} x_k y_k$. For curvature $-1/\kappa$ with $\kappa > 0$, the d-dimensional Lorentz model is defined as

$$\mathbb{H}^d_\kappa = \left\{ x \in \mathbb{R}^{d+1} \mid \langle x, x\rangle_{\mathcal{L}} = -\kappa,\ x_0 > 0 \right\}. \tag{1}$$

For $x, y \in \mathbb{H}^d_\kappa$, their geodesic distance is defined as

$$d_{\mathbb{H}}(x, y) = \sqrt{\kappa}\ \operatorname{arcosh}\left(-\frac{\langle x, y\rangle_{\mathcal{L}}}{\kappa} \right). \tag{2}$$

Hyperbolic Angle and Radial Depth Let $o = (\sqrt{\kappa}, 0, \ldots, 0)^\top$ denote origin. The radial depth of x is $r(x) = d_{\mathbb{H}}(o, x)$ and is commonly associated with relative depth in hyperbolic representations (Nickel & Kiela, 2017; 2018).

For $y, z \in \mathbb{H}^d_\kappa$ viewed from $x \in \mathbb{H}^d_\kappa$, their hyperbolic angle is

$$\angle_x(y, z) = \arccos\left(\frac{\left\langle \log^{\mathbb{H}}_x(y), \log^{\mathbb{H}}_x(z) \right\rangle_{\mathcal{L}}}{\left\| \log^{\mathbb{H}}_x(y) \right\|_{\mathcal{L}} \left\| \log^{\mathbb{H}}_x(z) \right\|_{\mathcal{L}}} \right). \tag{3}$$

Geodesic distance measures pairwise proximity, while radial depth and hyperbolic angle provide complementary ordering and directional cues.

Hyperbolic Entailment Cones Hyperbolic entailment cones model asymmetric and partially ordered relations through nested angular regions (Ganea et al., 2018a; Hasan et al., 2026). Each point $x \in \mathbb{H}^d_\kappa$ defines a cone $\mathcal{C}(x)$ with half-aperture $\psi(x)$. If $\theta_x(y)$ denotes the angle between the cone axis at x and the direction from x to y, then

$$\mathcal{C}(x) = \left\{ y \in \mathbb{H}^d_\kappa \mid \theta_x(y) \le \psi(x) \right\}. \tag{4}$$

Since the cone aperture varies with radial depth, cone inclusion is generally asymmetric, providing a natural geometric criterion for directed relations.

4 Proposed Method

Pretrained EEG tokens contain relational information that conventional pooling does not explicitly capture. RHOP (Hu et al., 2026) summarizes their first- and second-order statistics, but leaves sample-specific pairwise relations implicit. Motivated by the observed tree-like structure of pretrained token representations, we model these relations in hyperbolic space. Geodesic distance measures token proximity, but its symmetry cannot determine relation direction. We therefore combine geodesic similarity with entailment-cone compatibility, directional preference, and radial ordering to select and orient token pairs. The retained relations form a sample-specific weighted directed graph, which supports directional aggregation of pretrained token information. As illustrated in Fig. 2, the proposed HRAGP consists of three main stages: *hyperbolic token modeling*, *relation-aware graph induction*, and *directional graph readout*.

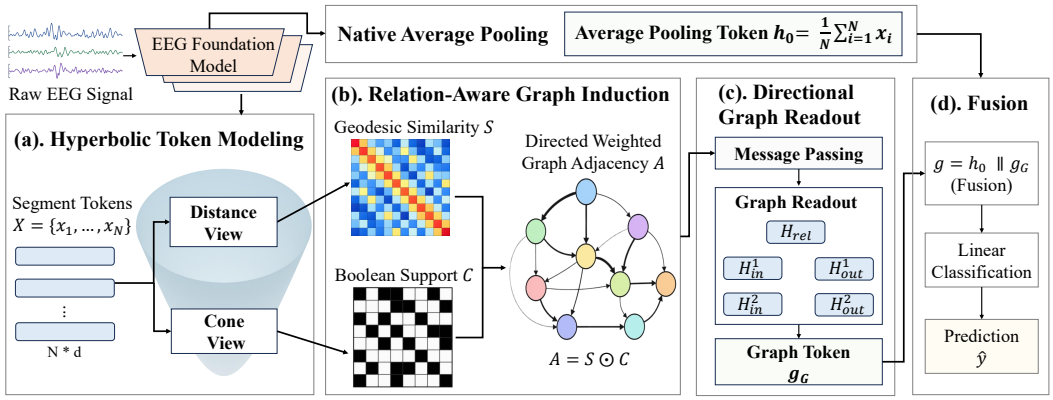


Figure 2: Overview of HRAGP. The segment tokens are modeled in hyperbolic space to induce a sample-specific directed graph, whose graph representation g_G is combined with the average-pooled backbone representation h_0 for downstream classification.

4.1 Hyperbolic Token Modeling

Given the segment tokens $X = \{x_i\}_{i=1}^{N}$ extracted from an EEG-FM, we first examine whether hyperbolic geometry is a useful inductive bias for their relations. The measurable tree-like structure of the token spaces motivates the use of hyperbolic geometry.

Because some EEG-FMs apply LayerNorm near their output, token normalization requires careful consideration in our geometric modeling. For geodesic similarity and radial ordering, variation in token magnitude helps distinguish tokens, whereas strong normalization may suppress this radial variation. For cone-based compatibility, however, the aperture of an entailment cone decreases as the radial norm of its root increases. Large or highly variable magnitudes can therefore make compatible token pairs difficult to retain. To accommodate these different requirements, we use two views of the same token sequence:

$$x_i^d = x_i, \qquad x_i^c = \mathrm{LN}(x_i),$$

where x_i^d is used for geodesic proximity and radial ordering, and x_i^c is used for angular compatibility.

To limit radial variation, token magnitudes are projected to a bounded interval. For branch $b \in \{d, c\}$, let $a_i^b = \|x_i^b\|_2$, with within-trial mean μ^b and standard deviation σ^b. We define

$$q_i^b = \frac{a_i^b - \mu^b}{\max(\sigma^b, \epsilon)}, \qquad \rho_i^b = \rho_{\min}^b + (\rho_{\max}^b - \rho_{\min}^b)\sigma(\alpha_b q_i^b), \tag{5}$$

where $\rho_{\min}^b$ and $\rho_{\max}^b$ specify the radial range.

For each view $b \in \{d, c\}$, we rescale x_i^b to obtain u_i^b with $\|u_i^b\|_2 = \rho_i^b$, and then map u_i^b to the Lorentz manifold:

$$z_i^b = \begin{bmatrix} \sqrt{\kappa + \|u_i^b\|_2^2} \\ u_i^b \end{bmatrix} \in \mathbb{H}_\kappa^d, \qquad b \in \{d, c\}. \tag{6}$$

4.2 Relation-Aware Graph Induction

Hyperbolic geometry offers complementary ways to characterize relations among tokens. Geodesic distance measures their intrinsic proximity on the manifold, whereas entailment cones provide an asymmetric criterion for directional compatibility. We use these geometric cues to construct a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ from the distance-view tokens $z_i^d i = 1^N$ and cone-view tokens $z_i^c i = 1^N$. Geodesic proximity determines edge weights, while cone compatibility and radial ordering determine which directed edges are retained.

Geodesic similarity. We first measure token similarity using geodesic distance, the length of the shortest path between two points on the hyperbolic manifold. Unlike Euclidean distance, it reflects both angular separation and radial position under the manifold geometry. For a pair of tokens (i, j), we compute the geodesic distance in the distance view and convert it into a continuous similarity,

$$S_{ij} = \exp\left(-\frac{d_{ij}}{\tau_d}\right), \tag{7}$$

where $d_{ij} = d_{\mathbb{H}}(z_i^d, z_j^d)$ is the pairwise geodesic distance, $\tau_d > 0$ controls the distance decay. Since $d_{ij} = d_{ji}$, the resulting similarity measures relation strength but does not specify its direction.

Directed relation support. Geodesic similarity assigns weights to nearby tokens, but its symmetry cannot determine direction and led to the information redundancy on relation weights . We design three complementary conditions: cone compatibility tests whether the candidate lies in a directed geometric region, directional preference resolves ambiguous bidirectional pairs, and radial ordering provides a shared orientation cue across the graph. For the cone view, let ψ_i denote the half-aperture of the entailment cone rooted at z_i^c, and let θ_{ij} denote the angle between its outward axis and the geodesic direction from z_i^c to z_j^c. We define the directional margin as $m_{ij} = \psi_i - \theta_{ij}$. In the distance view, we measure the radial depth of token i by $r_i = d_{\mathbb{H}}(o, z_i^d)$. A directed relation $i \to j$ is retained if the following three conditions hold:

$$\mathcal{R}_{ij} : \begin{cases} m_{ij} > \tau_{\text{margin}}, \\ m_{ij} - m_{ji} > \tau_{\text{pref}}, \\ r_j - r_i > \tau_{\text{depth}}. \end{cases} \tag{8}$$

1. **Cone compatibility.** The first condition requires z_j^c to lie inside the entailment cone rooted at z_i^c, with a sufficient angular margin. It excludes token pairs that are close in geodesic distance but do not satisfy the directed angular constraint. Specifically,

$$\psi_i = \arcsin\left(\frac{\lambda}{\sqrt{\kappa}\,\|(z_i^c)_{\text{space}}\|_2}\right), \quad \theta_{ij} = \arccos\left(\frac{\left\langle -\log_{z_i^c}^{\mathbb{H}}(o), \log_{z_i^c}^{\mathbb{H}}(z_j^c)\right\rangle_L}{\left\|\log_{z_i^c}^{\mathbb{H}}(o)\right\|_L \left\|\log_{z_i^c}^{\mathbb{H}}(z_j^c)\right\|_L}\right), \tag{9}$$

   where $\lambda > 0$ controls the cone aperture and κ is the curvature parameter of Lorentz model. The chosen cone-view radial range satisfies $\rho_{\min}^c > \lambda/\sqrt{\kappa}$, ensuring that the argument of $arcsin$ lies in $(0, 1)$.

2. **Directional preference.** Cone compatibility alone may not establish a sufficiently clear orientation for a candidate pair. The second condition retains $i \to j$ only when its directional margin exceeds that of $j \to i$ by at least τ_{pref}.

3. **Radial ordering.** The final condition imposes a shallow-to-deep orientation in the hyperbolic space. It complements the local angular constraints with a consistent radial order and prevents geometrically compatible pairs from being oriented arbitrarily.

The conjunction of the conditions gives the Boolean support $C_{ij} = \mathbb{I}[R_{ij}]$, where $\mathbb{I}[\cdot]$ equals one when all conditions in R_{ij} hold and zero otherwise. The weighted adjacency is then simply $A = S \odot C$, where C determines the existence and orientation of an edge and S determines its strength.

Permutation equivariance. The graph is independent of the token index. For any permutation matrix P, the adjacency satisfies $A(PX) = PA(X)P^\top$. Derivation is provided in App. B.6.

4.3 Directional Graph Readout

After graph induction, the graph contains both edge weights and orientation. We preserve its directionality during graph aggregation rather than symmetrizing the adjacency. For message aggregation, the distance view hyperbolic tokens are remapped to the tangent space at the Lorentz origin,

$$h_i = \left[\log_o^{\mathbb{H}}(z_i^d)\right]_{1:d}, \quad H = [h_1, \ldots, h_N]^\top. \tag{10}$$

Let $\widetilde{A}_{\mathrm{out}}$ denote outgoing adjacency obtained by one-sided square-root degree normalization of A, and define the incoming adjacency as $\widetilde{A}_{\mathrm{in}} = \widetilde{A}_{\mathrm{out}}^{\top}$. Directional propagation is given by

$$H_{\mathrm{out}}^{(\ell)} = \widetilde{A}_{\mathrm{out}} H_{\mathrm{out}}^{(\ell-1)}, \quad H_{\mathrm{in}}^{(\ell)} = \widetilde{A}_{\mathrm{in}} H_{\mathrm{in}}^{(\ell-1)}, \quad \ell = 1, 2, \tag{11}$$

with $H_{\mathrm{out}}^{(0)} = H_{\mathrm{in}}^{(0)} = H$. We also retain the neighbor relationship $H_{\mathrm{rel}} = H_{\mathrm{out}}^{(1)} - \mathrm{Diag}(\widetilde{A}_{\mathrm{out}}\mathbf{1})H$. The incoming and outgoing fields capture relational roles, while the second propagation step incorporates dual-hop dependencies. H_{rel} preserves the difference along outgoing relations. We apply mean pooling along the token dimension N to each relational field and concatenate the resulting summaries. Let $\bar{H}$ denote mean pooling over the token dimension, Final graph token g_G is given by

$$g_G = (\bar{H}_{\mathrm{out}}^{(1)}) \| \bar{H}_{\mathrm{in}}^{(1)} \| \bar{H}_{\mathrm{out}}^{(2)} \| \bar{H}_{\mathrm{in}}^{(2)} \| \bar{H}_{\mathrm{rel}}. \tag{12}$$

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate HRAGP on five EEG benchmarks covering distinct downstream paradigms: SEED-V (Liu et al., 2022) for emotion recognition, BCIC-2a (Brunner et al., 2008) for motor imagery, TUAB (Obeid & Picone, 2016) for abnormal EEG detection, TUEV (Obeid & Picone, 2016; Golmohammadi et al., 2019) for EEG event classification, and HMC (Alvarez-Estevez & Rijsman, 2022) for sleep staging. These datasets differ substantially in task type, recording protocol, and class structure, providing a heterogeneous evaluation setting.

Backbones. We choose four pretrained EEG-FMs: LaBraM (Jiang et al., 2024), EEGPT (Wang et al., 2024), CBraMod (Wang et al., 2025), and REVE (El Ouahidi et al., 2025). These models employ different tokenization schemes, pre-training objectives, and backbone designs, allowing us to examine whether HRAGP generalizes across heterogeneous pretrained representations.

Evaluation. All experiments follow the pipeline of EEG-FM-Bench (Xiong et al., 2025), including data preprocessing, pretrained checkpoints, dataset splits, and downstream optimization. We use the full-parameter single-task setting and replace only the original downstream readout with HRAGP. Results are reported as mean and standard deviation over five independent runs.

Baselines. We compare HRAGP with two downstream readouts for each EEG-FM. AvgPool is the conventional average-pooling readout used in EEG-FM-Bench (Xiong et al., 2025). RHOP (Hu et al., 2026) provides a structured comparison by summarizing first- and second-order token statistics. All three readouts use the same pretrained backbones, data partitions, preprocessing, and

Table 1: Comparison of downstream readouts on SEED-V and BCIC-2a.

Backbone	Readout	SEED-V			BCIC-2a		
		B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
LaBraM	AvgPool	24.35±0.55	23.64±1.21	5.61±1.13	32.37±0.21	30.79±0.39	9.83±0.30
	RHOP	39.76±0.36	36.70±1.21	24.35±0.57	35.21±0.89	34.11±0.99	13.61±1.15
	HRAGP	**42.66±0.55**	**38.11±1.90**	**27.60±0.77**	**37.63±0.86**	**34.13±2.00**	**16.84±1.19**
EEGPT	AvgPool	36.43±1.33	33.00±1.94	18.52±0.67	44.07±3.27	38.61±4.23	25.42±4.36
	RHOP	37.01±0.67	35.48±0.77	22.34±0.53	44.68±0.40	43.21±1.01	26.23±0.54
	HRAGP	**39.76±0.85**	**38.81±0.56**	**25.41±1.03**	**46.16±0.96**	**44.66±1.50**	**28.21±1.25**
CBraMod	AvgPool	28.34±0.17	27.48±0.89	10.55±0.20	33.71±0.86	28.99±1.35	11.61±1.14
	RHOP	**33.44±0.91**	**31.09±2.39**	**16.08±0.84**	32.34±1.43	31.25±1.47	9.79±1.99
	HRAGP	31.27±1.44	30.48±0.92	13.95±1.10	**37.57±0.77**	**35.71±0.74**	**16.76±1.02**
REVE	AvgPool	40.65±0.51	40.35±1.63	25.82±0.72	42.73±1.10	42.60±1.77	27.08±1.46
	RHOP	41.56±1.18	39.39±0.39	27.11±0.98	48.21±1.63	47.39±1.92	30.95±2.13
	HRAGP	**43.86±1.16**	**40.67±0.55**	**29.08±0.99**	**54.69±1.65**	**53.77±1.83**	**39.58±2.20**

Table 2: Comparison of downstream readouts on TUEV and TUAB.

Backbone	Readout	TUEV			TUAB		
		B-Acc	W-F1	Cohen's κ	B-Acc	AUROC	AUPRC
LaBraM	AvgPool	59.58±4.31	77.72±3.25	64.06±5.03	79.50±0.81	86.91±0.51	83.86±1.96
	RHOP	63.80±0.56	**84.20±0.38**	67.85±0.79	82.44±0.12	**91.05±0.11**	90.78±0.12
	HRAGP	**67.67±1.90**	84.03 ±1.51	**74.03±2.39**	**83.01±0.25**	90.44±0.48	**91.14±0.36**
EEGPT	AvgPool	**66.86±2.54**	83.95±0.14	74.45±0.09	78.66±0.24	87.52±1.01	86.94±2.15
	RHOP	64.28±1.52	86.70±0.43	78.58±0.80	**82.00±0.54**	**89.16±0.46**	**90.09±0.40**
	HRAGP	64.37±1.65	**87.80±0.56**	**80.44±1.21**	80.99±0.40	88.77±0.47	89.35±0.23
CBraMod	AvgPool	65.42±0.90	79.03±1.36	64.13±2.48	81.40±0.33	89.02±0.16	89.93±0.31
	RHOP	62.80±1.74	80.97±0.52	68.20±1.03	81.79±0.44	89.40±0.38	89.61±0.44
	HRAGP	**69.36±1.43**	**83.73±1.21**	**73.53±2.18**	**82.22±0.57**	**90.25±0.25**	**90.68±0.31**
REVE	AvgPool	73.98±1.21	90.25±1.42	85.07±1.80	81.32±0.35	87.93±1.76	87.38±1.56
	RHOP	77.02±1.29	93.32±0.27	89.49±0.41	81.69±0.88	89.56±0.30	89.56±0.51
	HRAGP	**78.46±1.88**	**94.41±0.61**	**91.31±1.19**	**82.17±0.16**	**89.63±0.52**	**89.70±0.67**

Table 3: Comparison of downstream readouts on HMC.

Readout	LaBraM			EEGPT		
	B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
AvgPool	70.85±0.44	71.52±0.55	64.47±0.70	69.67±1.24	73.03±0.32	64.64±0.92
RHOP	73.75±0.45	75.95±0.29	68.43±0.33	**74.62±0.30**	**76.33±0.26**	**69.22±0.14**
HRAGP	**75.36±0.59**	**77.62±0.37**	**70.52±0.36**	74.59±0.30	76.28±0.27	69.14±0.13
Readout	CBraMod			REVE		
	B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
AvgPool	71.14±0.14	72.76±0.55	64.86±0.41	71.82±0.82	75.08±0.74	66.49±0.83
RHOP	71.49±0.32	73.44±0.61	65.50±0.40	73.06±1.20	75.45±0.85	67.52±1.08
HRAGP	**72.43±0.56**	**74.55±0.41**	**66.74±0.43**	**74.27±0.63**	**75.73±0.25**	**68.09±0.45**

downstream training protocol. Further implementation are provided in Apps. D and D.4. More comparison with non-FM method is provided in App. E

Metrics. Balanced accuracy (B-Acc) is reported for all datasets. For multi-class tasks, we additionally report weighted F1 (W-F1) and Cohen's κ. For TUAB, AUROC and AUPRC are reported together with B-Acc. The same metrics and data partitions are used for all compared readouts.

5.2 Main Results

Overall performance. Tables 1–3 summarize the results across five downstream datasets and four EEG foundation models. HRAGP improves the corresponding AvgPool baseline in most comparisons, showing that the proposed relational readout transfers consistently across heterogeneous pretrained representations and EEG paradigms.

Comparison with RHOP. Compared with RHOP, HRAGP obtains higher mean performance in most of the comparisons. The advantage is most evident on BCIC-2a, where REVE+HRAGP exceeds RHOP by 6.48% B-Acc, 6.38% W-F1, and 8.63% Cohen's κ points. HRAGP also improves over RHOP mostly on SEED-V and TUEV. Two readouts are closer on TUAB and HMC, with RHOP remaining stronger for several EEGPT settings. This suggests that high-order distributional statistics and token relation modeling capture different aspects of pretrained EEG representations.

Parameter overhead. HRAGP introduces a small and stable overhead across backbones and datasets. The directional readout concatenates five d-dimensional relational summaries, yielding

a graph descriptor of approximately $5d$. Consequently, its additional parameters mainly come from the final readout and fusion layers and scale predictably with the backbone token dimension. In comparison, the parameter count of RHOP varies more noticeably across backbone and datasets. HRAGP remains comparable to RHOP in most settings and can be slightly larger in some datasets, while both are substantially smaller than the corresponding EEG-FM backbones. (see App. D.4) Since HRAGP does not change the backbone architecture, the same readout can be attached to backbones with different token dimensions with little architectural adjustment.

5.3 Ablation Studies

We conduct ablation studies on the key components of HRAGP: the geometric space for relation modeling, the geometric conditions used for graph induction, and the directional graph readout. Unless otherwise specified, other settings are kept identical to those in the main experiments.

Effect of hyperbolic relation modeling. We compare Euclidean and hyperbolic global similarity graphs to examine whether the improvement comes from graph aggregation or from the geometry used to model token relations. The Euclidean variant uses pairwise Euclidean distances between tokens, while the hyperbolic variant replaces them with geodesic distances after hyperbolic token modeling. Results and visualization is reported in Table 4 and Fig 3. The similarities are defined as

$$S_{ij}^{E} = \exp\left(-\frac{\|x_i - x_j\|_2}{\tau_E}\right), \qquad S_{ij}^{H} = \exp\left(-\frac{d_{\mathbb{H}}(z_i, z_j)}{\tau_d}\right). \tag{13}$$

Geometric relation conditions. We ablate the three geometric conditions used for directed graph induction in Table 5. On BCIC-2a, the full criterion achieves the best result on all three metrics, reaching 37.63% B-Acc, 34.13% W-F1, and 16.84% κ. Removing any condition degrades performance, with the largest drop observed when radial ordering is removed from the full configuration.

A similar trend is observed on TUEV, where the complete criterion achieves 67.67% B-Acc, 84.03% W-F1, and 74.03% κ. Individual conditions alone do not consistently outperform the unconstrained variant across all metrics, whereas their combination produces the strongest overall performance. It indicates that cone compatibility, directional preference, and radial ordering act as complementary constraints rather than independent similarity cues.

Effect of directional graph modeling. We examine whether preserving edge orientation is necessary for relation-aware graph readout. As shown in Table 6, we compare an undirected variant with incoming-only, outgoing-only, and full bidirectional propagation. Results showed that using either incoming or outgoing propagation alone yields similar performance, but both remain below the full bidirectional readout, indicating that the gain does not arise from either propagation direction alone; preserving both relational roles is important for exploiting the induced directed graph.

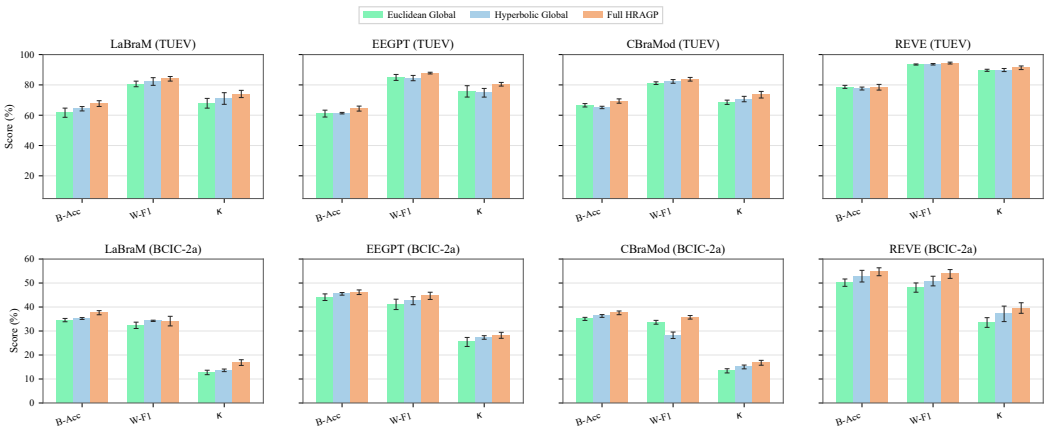


Figure 3: Visualization of the effect of hyperbolic relation modeling.

Table 4: Effect of hyperbolic relation modeling across different EEG foundation models. Euclidean Global uses a global similarity graph from pairwise Euclidean distances, whereas Hyperbolic Global uses geodesic distances. Full HRAGP denotes the complete proposed method.

Backbone	Variant	TUEV			BCIC2a		
		B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
LaBraM	Euclidean Global	61.66±3.03	80.64±1.85	67.89±3.21	34.55±0.69	32.37±1.32	12.73±0.93
	Hyperbolic Global	64.27±1.44	82.25±2.53	71.03±3.84	35.21±0.39	34.25±0.26	13.61±0.54
	Full HRAGP	**67.67±1.90**	**84.03±1.51**	**74.03±2.39**	**37.63±0.86**	**34.13±2.00**	**16.84±1.19**
EEGPT	Euclidean Global	61.07±2.27	84.94±1.97	75.73±3.75	44.09±1.38	41.10±2.16	25.44±1.88
	Hyperbolic Global	61.36±0.50	84.51±1.75	74.82±2.86	45.49±0.55	42.64±1.67	27.31±0.74
	Full HRAGP	64.37±1.65	**87.80±0.56**	**80.44±1.21**	**46.16±0.96**	**44.66±1.50**	**28.21±1.25**
CBraMod	Euclidean Global	66.56±1.12	81.15±0.89	68.55±1.41	35.05±0.64	33.64±0.80	13.40±0.87
	Hyperbolic Global	65.12±0.79	82.30±1.21	70.67±1.79	36.27±0.58	28.22±1.35	15.02±0.79
	Full HRAGP	**69.36±1.43**	**83.73±1.21**	**73.53±2.18**	**37.57±0.77**	**35.71±0.74**	**16.76±1.02**
REVE	Euclidean Global	**78.71±1.00**	93.47±0.37	89.64±0.69	50.15±1.53	48.10±1.93	33.54±2.02
	Hyperbolic Global	77.61±0.96	93.62±0.48	89.82±1.02	52.86±2.43	50.82±2.01	37.15±3.23
	Full HRAGP	78.46±1.88	**94.41±0.61**	**91.31±1.19**	**54.69±1.65**	**53.77±1.83**	**39.58±2.20**

Table 5: Ablation of conditions using LaBraM on BCIC-2a and TUEV. C_1: cone compatibility; C_2: directional preference; C_3: radial ordering.

C_1	C_2	C_3	BCIC-2a			TUEV		
			B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
×	✓	✓	36.32±0.73	33.11±0.80	15.09±0.99	65.60±2.35	83.50±1.57	72.96±3.20
✓	×	✓	36.88±0.98	33.94±1.20	15.89±1.31	66.50±1.71	82.50±1.78	71.37±2.56
✓	✓	×	35.45±0.83	32.64±0.99	13.93±1.09	67.11±1.22	83.61±0.84	73.26±1.59
×	×	✓	34.53±0.51	31.77±0.62	12.38±0.72	64.02±3.92	83.08±2.14	72.45±4.03
×	✓	×	34.61±0.86	32.46±0.78	12.85±1.15	66.96±1.90	82.91±1.33	71.31±1.73
✓	×	×	36.09±1.43	33.07±1.70	14.79±1.89	66.30±2.71	82.19±2.37	70.59±3.82
×	×	×	34.76±1.38	31.71±0.71	13.01±1.82	65.45±2.94	81.87±0.96	70.52±1.27
✓	✓	✓	**37.63±0.86**	**34.13±2.00**	**16.84±1.19**	**67.67±1.90**	**84.03±1.51**	**74.03±2.39**

Table 6: Ablation of edge orientation and directional readout with LaBraM on BCIC-2a and TUEV. Direction indicates whether edge orientation is preserved during graph induction; Out and In indicate aggregation along outgoing and incoming edges. ✓ and × denote enabled and disabled components, while − indicate readout branches that do not apply without directed edges.

Direction	Out.	In.	BCIC-2a			TUEV		
			B-Acc	W-F1	Cohen's κ	B-Acc	W-F1	Cohen's κ
×	–	–	35.58 ±1.29	32.33 ±1.90	13.84 ±1.69	66.79 ±1.01	83.03 ±2.07	72.19 ±3.12
✓	×	✓	35.59 ±1.09	33.09 ±1.19	14.12 ±1.44	66.57 ±1.52	82.53 ±1.74	71.13 ±1.74
✓	✓	×	35.56 ±1.53	32.74 ±1.46	14.07 ±2.03	67.30±2.49	82.28±2.68	71.21 ±4.26
✓	✓	✓	**37.63±0.86**	**34.13±2.00**	**16.84±1.19**	**67.67±1.90**	**84.03±1.51**	**74.03±2.39**

6 CONCLUSION

This work introduced HRAGP, a downstream readout for EEG foundation models that models sample-specific relations among pretrained EEG tokens. HRAGP combines hyperbolic proximity, entailment compatibility, and radial ordering to construct a weighted directed graph, which is summarized by directional graph readout and fused with the native average-pooled representation. Experiments in four EEG-FMs and five downstream benchmarks show that HRAGP improves conventional average pooling and recent work RHOP in nearly all comparisons. The ablation results support the roles of hyperbolic relation modeling, geometric relation constraints, and directional propagation. These results suggest that modeling token-level relations can provide complementary information to conventional global pooling for downstream EEG-FM adaptation.

AI USE STATEMENT

In this work, we used generative AI tools to assist with manuscript editing for readability. We reviewed all AI-assisted content, and AI-assisted code was manually inspected and tested for correctness. We take responsibility for the final content of this work, including all text, mathematical expressions, references, and software artifacts produced with the aid of generative AI.

REFERENCES

Diego Alvarez-Estevez and Roselyne Rijsman. Haaglanden Medisch Centrum sleep staging database. *PhysioNet*, March 2022. doi: 10.13026/t79q-fr32. Version 1.1.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for spd neural networks. *Advances in neural information processing systems*, 32, 2019.

Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. BCI competition 2008–graz data set a. *Institute for knowledge discovery (laboratory of brain-computer interfaces), Graz University of Technology*, 16(1-6):1, 2008.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32, 2019.

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 D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 22541–22577. Curran Associates, Inc., 2025. doi: 10.52202/085713-0760.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings. In *International conference on machine learning*, pp. 1646–1655. PMLR, 2018a.

Octavian Ganea, Gary Becigneul, and Thomas Hofmann. Hyperbolic neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018b.

Meysam Golmohammadi, Amir Hossein Harati Nejad Torbati, Silvia Lopez de Diego, Iyad Obeid, and Joseph Picone. Automatic analysis of eegs using big data and hybrid deep learning architectures. *Frontiers in Human Neuroscience*, Volume 13 - 2019, 2019. ISSN 1662-5161. doi: 10.3389/fnhum.2019.00076.

Mikhael Gromov. Hyperbolic groups. In *Essays in group theory*, pp. 75–263. Springer, 1987.

Zahid Hasan, Masud Ahmed, and Nirmalya Roy. Lorentz entailment cone for semantic segmentation. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5216–5225. IEEE, 2026.

Chen Hu, Rui Wang, Xiaoning Song, Tao Zhou, Xiao-Jun Wu, Nicu Sebe, and Ziheng Chen. A correlation manifold self-attention network for eeg decoding. In *Proceedings of International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 5372–5380, 2025.

Chen Hu, Ziheng Chen, Rui Wang, Yefeng Zheng, and Nicu Sebe. Riemannian high-order pooling for brain foundation models. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=66h1sCMm7F.

Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Wei-Bang Jiang, Liming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *International Conference on Learning Representations*, volume 2024, pp. 16405–16426, 2024.

Demetres Kostas, Stéphane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, Volume 15 - 2021, 2021. ISSN 1662-5161. doi: 10.3389/fnhum.2021.653659.

Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, and Yogatheesan Varatharajah. Eeg foundation models: a critical review of current progress and future directions. *Journal of neural engineering*, 23(2):021001, 2026.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Shanglin Li, Chu Shiwen, Okan Koç, Yi Ding, Qibin Zhao, Motoaki Kawanabe, and Ziheng Chen. Heegnet: Hyperbolic embeddings for eeg. In *International Conference on Learning Representations*, volume 2026, pp. 74317–74346, 2026.

Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, and Dongrui Wu. Eeg-fm-compass: Progress, benchmarking, and future directions for eeg foundation models. *National Science Review*, pp. nwag466, 2026.

Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. *Advances in neural information processing systems*, 32, 2019.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2):715–729, 2022. doi: 10.1109/TCDS.2021.3071170.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3779–3788. PMLR, 10–15 Jul 2018.

Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in neuroscience*, 10:196, 2016.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 31116–31129. Curran Associates, Inc., 2022. doi: 10.52202/068431-2256.

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

Guagnyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of eeg signals. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 39249–39280. Curran Associates, Inc., 2024. doi: 10.52202/079017-1239.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. In *International conference on learning representations*, volume 2025, pp. 75310–75346, 2025.

Wei Xiong, Jiangtong Li, Jie Li, and Kun Zhu. Eeg-fm-bench: A comprehensive benchmark for the systematic evaluation of eeg foundation models, 2025. URL https://arxiv.org/abs/2508.17742.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36:78240–78260, 2023.

A NOTATIONS

For convenience, Table 7 summarizes the main notation used throughout the paper.

Table 7: Notations used in this paper.

Symbol	Meaning
$X = \{x_i\}_{i=1}^{N}$	Segment-token sequence produced by an EEG foundation model
$N,\ d$	Number of tokens and token feature dimension
$x_i^d,\ x_i^c$	Distance-view and cone-view tokens
$a_i^b,\ q_i^b,\ \rho_i^b$	Token magnitude, standardized magnitude, and bounded radial magnitude, $b \in \{d, c\}$
$u_i^b,\ z_i^b$	Spatial component and corresponding Lorentz representation
$\mathbb{H}_\kappa^d$	d-dimensional Lorentz model with curvature $-1/\kappa$
$\langle\cdot,\cdot\rangle_{\mathcal{L}}$	Lorentzian inner product
$d_{\mathcal{H}}(x, y)$	Hyperbolic geodesic distance between x and y
$o,\ r_i$	Lorentz origin and radial depth of token i
$\psi_i,\ \theta_{ij}$	Entailment-cone half-aperture and angular deviation
m_{ij}	Directional cone margin, $m_{ij} = \psi_i - \theta_{ij}$
S_{ij}	Geodesic similarity between tokens i and j
C_{ij}	Boolean support of the directed relation $i \to j$
A_{ij}	Weighted directed adjacency, $A_{ij} = C_{ij}S_{ij}$
$\mathcal{G} = (V, E, A)$	Sample-specific weighted directed relation graph
$H_{\text{out}}^{(\ell)},\ H_{\text{in}}^{(\ell)}$	Token features after ℓ outgoing and incoming propagation steps
H_{rel}	Residual relational field
g_G	Graph-level relational representation produced by HRAGP
h_0	Native average-pooled token representation
$g = h_0 \| g_G$	Fused representation used for downstream classification
$\tau_d,\ \tau_{\text{margin}},\ \tau_{\text{pref}},\ \tau_{\text{depth}}$	Distance, cone-margin, directional-preference, and radial-order thresholds

B PRELIMINARIES

This section provides the geometric operations and implementation details omitted from the main text for brevity. We follow the Lorentz convention used throughout the paper, where $\mathbb{H}_\kappa^d$ has constant curvature $-1/\kappa$ with $\kappa > 0$.

B.1 LORENTZ MODEL AND BASIC OPERATIONS

For $x, y \in \mathbb{R}^{d+1}$, the Lorentzian inner product is

$$\langle x, y\rangle_{\mathcal{L}} = -x_0 y_0 + \sum_{k=1}^{d} x_k y_k. \tag{14}$$

The d-dimensional Lorentz model is

$$\mathbb{H}_\kappa^d = \left\{ x \in \mathbb{R}^{d+1} \,\middle|\, \langle x, x\rangle_{\mathcal{L}} = -\kappa,\ x_0 > 0 \right\}. \tag{15}$$

For a tangent vector $v \in T_x\mathbb{H}_\kappa^d$, $\langle x, v\rangle_{\mathcal{L}} = 0$, and its Lorentz norm is

$$\|v\|_{\mathcal{L}} = \sqrt{\langle v, v\rangle_{\mathcal{L}}}. \tag{16}$$

For $x, y \in \mathbb{H}_\kappa^d$, define

$$\alpha(x, y) = -\frac{\langle x, y\rangle_{\mathcal{L}}}{\kappa}. \tag{17}$$

Since $\alpha(x, y) \geq 1$, their geodesic distance is

$$d_{\mathcal{H}}(x, y) = \sqrt{\kappa}\ \operatorname{arcosh}(\alpha(x, y)). \tag{18}$$

The canonical origin is

$$o = (\sqrt{\kappa}, 0, \ldots, 0)^\top. \tag{19}$$

B.2 Exponential and Logarithmic Maps

For $v \in T_x\mathbb{H}^d_\kappa$, the exponential map is

$$\mathrm{Exp}^{\mathcal{H}}_x(v) = \cosh\left(\frac{\|v\|_{\mathcal{L}}}{\sqrt{\kappa}}\right) x + \sqrt{\kappa}\,\sinh\left(\frac{\|v\|_{\mathcal{L}}}{\sqrt{\kappa}}\right)\frac{v}{\|v\|_{\mathcal{L}}}, \tag{20}$$

where the zero-vector case is defined by continuity as $\mathrm{Exp}^{\mathcal{H}}_x(0) = x$.

Conversely, for $x, y \in \mathbb{H}^d_\kappa$, let $\alpha = \alpha(x, y)$. The logarithmic map from y to the tangent space at x is

$$\mathrm{Log}^{\mathcal{H}}_x(y) = \frac{\mathrm{arcosh}(\alpha)}{\sqrt{\alpha^2 - 1}}\,(y - \alpha x)\,. \tag{21}$$

Equivalently, $\|\,\mathrm{Log}^{\mathcal{H}}_x(y)\|_{\mathcal{L}} = d_{\mathcal{H}}(x, y)$.

In HRAGP, the logarithmic map is used only after graph induction. Specifically, the distance-view representation z^d_i is mapped to $T_o\mathbb{H}^d_\kappa$ and its spatial coordinates are used as the Euclidean message-passing feature,

$$h_i = \left[\mathrm{Log}^{\mathcal{H}}_o(z^d_i)\right]_{1:d}. \tag{22}$$

B.3 Hyperbolic Angle and Radial Depth

Given $x, y, z \in \mathbb{H}^d_\kappa$, the angle between the geodesics from x to y and z is evaluated in the tangent space $T_x\mathbb{H}^d_\kappa$:

$$\angle_x(y, z) = \arccos \frac{\left\langle \mathrm{Log}^{\mathcal{H}}_x(y), \mathrm{Log}^{\mathcal{H}}_x(z)\right\rangle_{\mathcal{L}}}{\left\|\mathrm{Log}^{\mathcal{H}}_x(y)\right\|_{\mathcal{L}}\left\|\mathrm{Log}^{\mathcal{H}}_x(z)\right\|_{\mathcal{L}}}. \tag{23}$$

The radial depth of x is its distance from the Lorentz origin,

$$r(x) = d_{\mathcal{H}}(o, x). \tag{24}$$

For the distance-view tokens, we use the shorthand $r_i = r(z^d_i)$. In HRAGP, radial depth is not used as an independent feature; it provides an ordering cue when determining the orientation of candidate token relations.

B.4 Hyperbolic Entailment Cones

For a point $x \in \mathbb{H}^d_\kappa$, we define an entailment cone rooted at x by

$$\mathcal{C}(x) = \left\{ y \in \mathbb{H}^d_\kappa \,\middle|\, \theta_x(y) \le \psi(x) \right\}, \tag{25}$$

where $\theta_x(y)$ is the angular deviation from the cone axis and $\psi(x)$ is the cone half-aperture.

Following Lorentz entailment-cone constructions, the half-aperture is parameterized using the spatial component of the Lorentz point:

$$\psi(x) = \arcsin\left(\frac{\lambda}{\sqrt{\kappa}\,\|x_{\mathrm{space}}\|_2}\right), \tag{26}$$

where κ is the squared hyperboloid radius, corresponding to constant curvature $-1/\kappa$,λ controls the cone scale.In this work, we set $\lambda = 1$. Since $\|(z^c_i)_{\mathrm{space}}\|_2 = \rho^c_i \ge \rho^c_{\min} > \lambda/\sqrt{\kappa}$, we have $0 < \lambda/(\sqrt{\kappa}\rho^c_i) < 1$.

B.5 Hyperbolic Token Mapping

Given a pretrained token x_i, HRAGP constructs the distance and cone views as

$$x^d_i = x_i, \qquad x^c_i = \mathrm{LN}(x_i). \tag{27}$$

For branch $b \in \{d, c\}$, let $a^b_i = \|x^b_i\|_2$. We standardize the token magnitude within each sample,

$$q^b_i = \frac{a^b_i - \mu^b}{\max(\sigma^b, \epsilon)}, \tag{28}$$

where μ^b and σ^b are the sample-wise mean and standard deviation of $\{a_i^b\}_{i=1}^N$.

The standardized magnitude is then mapped to a bounded radial range,

$$\rho_i^b = \rho_{\min}^b + \left(\rho_{\max}^b - \rho_{\min}^b\right) \sigma\left(\alpha_b q_i^b\right), \tag{29}$$

where $\sigma(\cdot)$ is the sigmoid function. After assigning the spatial component u_i^b the prescribed magnitude $\|u_i^b\|_2 = \rho_i^b$, the corresponding Lorentz point is obtained by

$$z_i^b = \begin{bmatrix} \sqrt{\kappa + \|u_i^b\|_2^2} \\ u_i^b \end{bmatrix}, \qquad b \in \{d, c\}. \tag{30}$$

By construction,

$$\left\langle z_i^b, z_i^b \right\rangle_{\mathcal{L}} = -\kappa, \tag{31}$$

and hence $z_i^b \in \mathbb{H}_\kappa^d$.

The bounded radial mapping prevents extreme token magnitudes from dominating hyperbolic depth while preserving sample-wise magnitude ordering. The distance view is used for geodesic proximity and radial ordering, whereas the cone view is used for angular compatibility.

B.6 Permutation Properties of HRAGP

We finally provide the permutation properties stated in the main text.

Proposition 1. Let $P \in \{0,1\}^{N\times N}$ be a permutation matrix acting on the token dimension. The graph induction operator of HRAGP is permutation equivariant:

$$A(PX) = PA(X)P^\top. \tag{32}$$

Proof. The two token views are constructed independently for each token, while the sample-wise magnitude statistics are invariant to token permutation. Therefore their Lorentz representations are permuted in the same manner as the input tokens. Pairwise geodesic distances, angular margins, and radial-depth comparisons are functions only of the corresponding token pair and are thus re-indexed by the same permutation. Consequently,

$$S(PX) = PS(X)P^\top, \qquad C(PX) = PC(X)P^\top. \tag{33}$$

Since the weighted adjacency is obtained element-wise as $A = C \odot S$,

$$A(PX) = \left(PCP^\top\right) \odot \left(PSP^\top\right) = P(C \odot S)P^\top, \tag{34}$$

□

Proposition 2. The graph-level representation produced by the directional readout is permutation invariant:

$$g_G(PX) = g_G(X). \tag{35}$$

Proof. Under a token permutation, the tangent-space feature matrix and the normalized incoming and outgoing adjacency matrices transform as

$$H' = PH, \qquad \widetilde{A}'_{\text{out}} = P\widetilde{A}_{\text{out}}P^\top, \qquad \widetilde{A}'_{\text{in}} = P\widetilde{A}_{\text{in}}P^\top. \tag{36}$$

It follows recursively that

$$H'^{(\ell)}_{\text{out}} = PH^{(\ell)}_{\text{out}}, \qquad H'^{(\ell)}_{\text{in}} = PH^{(\ell)}_{\text{in}}, \tag{37}$$

and the same equivariance holds for H_{rel}. Mean pooling over the token dimension is invariant to row permutation, so all five pooled relational summaries remain unchanged. Their concatenation therefore satisfies $g_G(PX) = g_G(X)$. □

Table 8: Details of the downstream EEG datasets used in our experiments. The numbers of training, validation, and test samples follow the EEG-FM-Bench processing protocol.

Dataset	Task	Channels	Duration (s)	Train	Valid	Test	Classes
TUAB	Abnormal EEG detection	23	30	247,728	12,315	12,277	2
TUEV	EEG event classification	21	5	87,834	12,473	13,046	6
SEED-V	Emotion recognition	60	10	3,552	4,638	4,128	5
HMC	Sleep staging	4	30	91,681	22,804	22,440	5
BCIC-2a	Motor imagery	22	4	2,784	1,152	1,152	4

B.7 Gromov Hyperbolicity Estimation

We estimate the tree-like structure of pretrained EEG token spaces using the standard four-point definition of Gromov hyperbolicity. Let $(\mathcal{X}, d)$ be a finite metric space. For any four points $x_i, x_j, x_k, x_l \in \mathcal{X}$, define the three pairwise sums

$$\begin{aligned} s_1 &= d(x_i, x_j) + d(x_k, x_l), \\ s_2 &= d(x_i, x_k) + d(x_j, x_l), \\ s_3 &= d(x_i, x_l) + d(x_j, x_k). \end{aligned} \tag{38}$$

Let $s_{(1)} \geq s_{(2)} \geq s_{(3)}$ denote these three values after sorting in descending order. The four-point hyperbolicity of the quadruple is

$$\delta(x_i, x_j, x_k, x_l) = \frac{1}{2}\left(s_{(1)} - s_{(2)}\right). \tag{39}$$

The Gromov hyperbolicity constant of the finite token space is then estimated by

$$\delta = \max_{i,j,k,l} \delta(x_i, x_j, x_k, x_l). \tag{40}$$

Since the absolute value of δ depends on the scale of the metric space, we report the normalized hyperbolicity

$$\delta_{\text{norm}} = \frac{\delta}{D}, \qquad D = \max_{i,j} d(x_i, x_j), \tag{41}$$

where D is the diameter of the token space. A smaller δ_{norm} indicates a metric space that is closer to a tree metric, while larger values indicate weaker tree-like structure.

For each pretrained backbone, the same procedure is applied separately to the raw token representations and their LayerNorm-normalized counterparts. This analysis is used only as an empirical diagnostic of the representation geometry and does not assume that the token space is intrinsically hyperbolic.

C Dataset and Preprocessing Details

We follow the dataset processing and evaluation protocol of EEG-FM-Bench (Xiong et al., 2025) to ensure consistent comparisons across EEG foundation models. This section summarizes the five downstream datasets used in our experiments and the corresponding preprocessing and partition protocols.

C.1 Dataset Overview

Our evaluation covers five representative EEG classification paradigms: abnormal EEG detection, EEG event classification, emotion recognition, sleep staging, and motor imagery. Table 8 summarizes the processed datasets used for downstream evaluation.

TUAB. The Temple University Hospital Abnormal EEG Corpus (TUAB) is used for binary abnormal EEG detection. Recordings are divided into normal and abnormal categories, providing a clinically oriented benchmark for evaluating whether pretrained representations transfer to long-duration EEG classification.

TUEV. The Temple University Hospital EEG Events Corpus (TUEV) contains annotated EEG events and is formulated as a six-class event classification task. Compared with recording-level classification, TUEV requires discrimination between shorter and often highly imbalanced event patterns.

SEED-V. SEED-V is an emotion-recognition dataset containing EEG recordings associated with five emotional states. It provides a substantially different downstream setting from the clinical datasets and is used to evaluate transfer to affective EEG representations.

HMC. The Haaglanden Medisch Centrum (HMC) sleep dataset is used for five-stage sleep classification. Continuous sleep recordings are segmented into 30-second epochs, following the conventional temporal resolution for sleep staging.

BCIC-2a. BCI Competition IV Dataset 2a (BCIC-2a) is a four-class motor-imagery benchmark containing left-hand, right-hand, feet, and tongue imagery. It is included to evaluate HRAGP on cross-subject sensorimotor EEG decoding.

C.2 Evaluation Partitions

We retain the dataset partitions defined by EEG-FM-Bench rather than constructing new train–test splits. The partition strategies account for the acquisition protocols and class distributions of individual datasets.

For TUAB, the original training set is retained, while the original evaluation set is divided by subject into equally sized validation and test subsets. TUEV is repartitioned using stratified sampling because of its strongly imbalanced event distribution, with approximately 80%, 10%, and 10% of the processed samples assigned to the training, validation, and test sets, respectively.

For SEED-V, the 15 experimental trials are divided into training, validation, and test groups using a 1:1:1 trial-level split, following the benchmark protocol. BCIC-2a uses a subject-level split: subjects 1–5 are used for training, subjects 6–7 for validation, and subjects 8–9 for testing. For HMC, subjects are randomly partitioned into training, validation, and test groups with a ratio of $103{:}24{:}24$. These partitions are kept identical for AvgPool, RHOP, and HRAGP in all comparisons.

C.3 Preprocessing Protocol

We use the standardized preprocessing pipeline provided by EEG-FM-Bench and preserve model-specific input conventions required by the released pretrained checkpoints. The raw recordings are first resampled to the target sampling rate expected by each backbone. Low-frequency drift is suppressed using an FIR-based filtering procedure, while power-line interference is removed with a 50 or 60 Hz notch filter according to the recording source. Unsupported channels are discarded, and the remaining electrodes are mapped by channel name to the standard 10–10 montage. Signal units are standardized before being passed to the model-specific adapters.

Continuous recordings are then divided into fixed-duration segments according to the downstream task, as summarized in Table 8. The same processed segments and data partitions are shared by all readout methods, ensuring that differences between AvgPool, RHOP, and HRAGP arise from the downstream readout rather than dataset processing.

Because pretrained EEG foundation models were developed with different input conventions, EEG-FM-Bench further preserves checkpoint compatibility through backbone-specific adapters. EEGPT follows the amplitude convention of its released implementation, whereas LaBraM and CBraMod use the corresponding $100\,\mu$V-style scaling. For LaBraM and CBraMod, the spatial channel layout is flattened into the one-dimensional channel ordering expected by the pretrained models. REVE uses its released positional encoding and checkpoint-compatible preprocessing; following EEG-FM-Bench, window-level normalization is used rather than statistics estimated from an entire test session to avoid information leakage.

All preprocessing, dataset partitions, and pretrained checkpoints are therefore held fixed across readout variants. HRAGP operates only on the segment-token representations produced by the corresponding pretrained backbone and does not change the backbone architecture.

Table 9: Common downstream fine-tuning configuration.

Configuration	Setting
Learning Rate	1e-6,5e-4
Optimization	AdamW
Training epochs	10-30
Warm-up epochs	3-5
Batch size	128
Gradient clipping	3.0
Default weight decay	0.01
Scheduler	Warm-up + cosine annealing
EEGPT scheduler	OneCycle
Number of runs	5

D EXPERIMENTAL CONFIGURATION

All experiments are conducted within the EEG-FM-Bench framework and follow its single-task full-parameter fine-tuning protocol unless otherwise specified. We keep the preprocessing, dataset splits, pretrained checkpoints, and optimization protocol fixed across different downstream readouts so that the comparison reflects only the effect of the readout module.

D.1 TRAINING SETUP

We use AdamW for downstream optimization. Following example of EEG-FM-Bench, models are trained for 10 or 30 epochs according to the different dataset, including 3-5 warm-up epochs. A two-stage learning-rate schedule consisting of linear warm-up followed by cosine annealing is used for most backbones, while EEGPT follows the OneCycle schedule used in its original implementation. The batch size is set to 128 and gradient clipping is applied with a maximum norm of 3.0.

The default weight decay is 0.01, while CBraMod uses 0.05 following the benchmark configuration. Backbone and readout learning rates are allowed to differ, with the backbone learning rate set to a fraction of the downstream-head learning rate. The final learning-rate combination for each backbone–dataset pair is selected according to validation-set performance. Each experiment is repeated with five random seeds, and the mean and standard deviation are reported.

Table 9 summarizes the common training configuration.

D.2 HRAGP CONFIGURATION

HRAGP introduces a small set of geometric and graph-construction hyperparameters in addition to the backbone fine-tuning configuration. The Lorentz manifold is parameterized by the hyperboloid radius κ, corresponding to constant curvature $-1/\kappa$. The distance and cone views use bounded radial ranges $[\rho^d_{\min}, \rho^d_{\max}]$ and $[\rho^c_{\min}, \rho^c_{\max}]$, respectively.

The weighted graph is controlled by four quantities: τ_d determines the decay of geodesic similarity, τ_{margin} controls cone compatibility, τ_{pref} controls directional preference, and τ_{depth} controls radial ordering. The directional readout uses two propagation steps, as defined in the main text.

Unless otherwise stated, the same HRAGP formulation and parameterization are used across backbones. Any backbone-specific differences arise only from the dimensionality and tokenization scheme of the pretrained EEG-FM.

D.3 MODEL-SPECIFIC INPUT ADAPTATION

We retain the backbone-specific input conventions of EEG-FM-Bench to preserve compatibility with the released pretrained checkpoints. EEGPT removes unsupported electrodes through its adapter, while LaBraM and CBraMod flatten the spatial channel layout into the one-dimensional ordering expected by their original implementations. REVE uses its released positional encoding and checkpoint-compatible preprocessing configuration.

Table 10: Comparison of downstream representation width and trainable parameter overhead between HRAGP and RHOP. D denotes the backbone token dimension, C the number of downstream classes, and T the temporal dimension used by RHOP. $\Delta = P_{\mathrm{HRAGP}} - P_{\mathrm{RHOP}}$; negative values indicate that HRAGP introduces fewer trainable parameters.

Dataset	Backbone	D	C	T	Feature width		Trainable parameters		
					HRAGP	RHOP	HRAGP	RHOP	Δ
TUEV	LaBraM	200	6	5	1,200	980	7.607K	5.886K	+1.721K
	CBraMod	200	6	5	1,200	980	7.607K	5.886K	+1.721K
	EEGPT	512	6	5	3,072	1,292	19.463K	7.758K	+11.705K
	REVE	512	6	5	3,072	1,292	19.463K	7.758K	+11.705K
TUAB	LaBraM	200	2	10	1,200	4,665	2.803K	9.332K	-6.529K
	CBraMod	200	2	10	1,200	4,665	2.803K	9.332K	-6.529K
	EEGPT	512	2	10	3,072	4,977	7.171K	9.956K	-2.785K
	REVE	512	2	10	3,072	4,977	7.171K	9.956K	-2.785K
SEED-V	LaBraM	200	5	10	1,200	4,665	6.406K	23.330K	-16.924K
	CBraMod	200	5	10	1,200	4,665	6.406K	23.330K	-16.924K
	EEGPT	512	5	10	3,072	4,977	16.390K	24.890K	-8.500K
	REVE	512	5	10	3,072	4,977	16.390K	24.890K	-8.500K
HMC	LaBraM	200	5	10	1,200	4,665	6.406K	23.330K	-16.924K
	CBraMod	200	5	10	1,200	4,665	6.406K	23.330K	-16.924K
	EEGPT	512	5	10	3,072	4,977	16.390K	24.890K	-8.500K
	REVE	512	5	10	3,072	4,977	16.390K	24.890K	-8.500K
BCIC-2a	LaBraM	200	4	4	1,200	696	5.205K	2.788K	+2.417K
	CBraMod	200	4	4	1,200	696	5.205K	2.788K	+2.417K
	EEGPT	512	4	4	3,072	1,008	13.317K	4.036K	+9.281K
	REVE	512	4	4	3,072	1,008	13.317K	4.036K	+9.281K

These adaptations are applied identically to all downstream readouts. HRAGP operates only on the segment-token sequence produced by the backbone and therefore does not modify the pretrained encoder or its input adapter.

D.4 Parameter Analysis and Overhead

We further compare the downstream parameter overhead of HRAGP and RHOP. Since the two methods construct representations in different ways, we report both the dimensionality of the fused representation and the actual number of trainable parameters introduced in the downstream path.

For HRAGP, the graph readout concatenates five D-dimensional relational summaries and fuses them with the native D-dimensional average-pooled representation. The resulting feature width is therefore $6D$. The trainable parameters introduced by the HRAGP readout and classifier are

$$P_{\mathrm{HRAGP}} = 2D + 1 + 6DC + C, \tag{42}$$

where D is the backbone token dimension and C is the number of downstream classes. 1 comes from learnable hyperparameter g_{gate} which adjust the raw value of final graph token g_G.

For RHOP, following its original construction, let

$$n = T + 3, \qquad q = \frac{n(n+1)}{2} + 3, \qquad g = \frac{q(q+1)}{2}, \tag{43}$$

where T denotes the number of temporal tokens involved in its high-order representation. The fused feature width is $D + g$, and the corresponding classifier contains

$$P_{\mathrm{RHOP}} = C(D + g) + C \tag{44}$$

trainable parameters.

Table 10 reports the resulting representation widths and actual trainable parameter counts for all backbone–dataset combinations used in our experiments.

The comparison reveals different scaling behaviors between the two readouts. HRAGP has a fixed fused width of $6D$ and its parameter count grows linearly with the backbone dimension D and the

Table 11: Additional comparison on TUEV and TUAB. Values are mean ± standard deviation in percent.

Backbone	Readout	TUEV			TUAB		
		B-Acc	W-F1	Kappa	B-Acc	AUROC	AUPRC
LaBraM	Base	64.09±0.65	83.12±0.52	66.37±0.93	81.40±0.19	90.22±0.09	89.65±0.16
	iSQRT-COV	62.36±2.26	80.62±0.94	61.47±2.34	81.88±0.23	90.60±0.12	90.39±0.18
	SVD-Padé	56.05±2.17	79.00±1.01	57.98±2.89	82.02±0.17	90.72±0.11	90.62±0.14
	iSICE	64.05±2.39	81.82±0.91	61.34±3.02	81.83±0.18	90.59±0.16	90.37±0.16
	RHOP	63.80±0.56	84.20±0.38	67.85±0.79	82.44±0.12	91.05±0.11	90.78±0.12
	AvgPool	59.58±4.31	77.72±3.25	64.06±5.03	79.50±0.81	86.91±0.51	83.86±1.96
	HRAGP	67.67±1.90	84.03±1.51	74.03±2.39	83.01±0.25	90.44±0.48	91.14±0.36
EEGPT	AvgPool	66.86±2.54	83.95±0.14	74.45±0.09	78.66±0.24	87.52±1.01	86.94±2.15
	HRAGP	64.37±1.65	87.80±0.56	80.44±1.21	80.99±0.40	88.77±0.47	89.35±0.23
CBraMod	AvgPool	65.42±0.90	79.03±1.36	64.13±2.48	81.40±0.33	89.02±0.16	89.93±0.31
	HRAGP	69.36±1.43	83.73±1.21	73.53±2.18	82.22±0.57	90.25±0.25	90.68±0.31
REVE	AvgPool	73.98±1.21	90.25±1.42	85.07±1.80	81.32±0.35	87.93±1.76	87.38±1.56
	HRAGP	78.46±1.88	94.41±0.61	91.31±1.19	82.17±0.16	89.63±0.52	89.70±0.67

number of classes C. It is therefore independent of the temporal token count used to construct the relational representation.

In contrast, the dimensionality of RHOP depends on T through two successive upper-triangular vectorizations in Eq. 43. Consequently, its representation dimension and classifier size increase nonlinearly with T. This difference is particularly visible on TUAB, SEED-V, and HMC, where $T = 10$ and HRAGP requires fewer trainable downstream parameters than RHOP. For example, with LaBraM on SEED-V and HMC, the trainable overhead decreases from 23.330K for RHOP to 6.406K for HRAGP.

On datasets with smaller T, such as TUEV and BCIC-2a, RHOP can produce a narrower downstream representation and therefore use fewer parameters. HRAGP nevertheless retains a predictable parameter budget: approximately 2.8–7.6K parameters for 200-dimensional backbones and 7.2–19.5K parameters for 512-dimensional backbones across the evaluated tasks. These quantities remain small compared with the pretrained EEG foundation models and require no modification of the backbone itself.

E Additional Comparison

E.1 Additional Comparison with Published Pooling Baselines

Table 11 places our results alongside the LaBraM-Base results reported by RHOP (Hu et al., 2026), including three global covariance-pooling (GCP) heads. The published RHOP results use its original evaluation protocol and three runs, whereas our results follow EEG-FM-Bench and use five runs. The two groups therefore provide broader context, not a protocol-matched comparison.

Table 12 places HRAGP alongside EEG models trained for individual downstream tasks without foundation-model pretraining. HRAGP with REVE reports higher balanced accuracy than the listed task-specific models on all five datasets. The results also depend on the backbone: for example, HRAGP with LaBraM performs well on TUEV, TUAB, SEED-V, and HMC, but trails several task-specific models on BCIC-2a. Because these results come from different evaluation protocols, the table provides broader context; the within-backbone comparisons in the main text assess the contribution of the HRAGP readout.

F Additional Ablation Studies

We provide additional ablation studies to further examine several design choices of HRAGP that are omitted from the main paper for brevity. Unless otherwise specified, all experiments in this section are conducted on TUAB using the LaBraM-Base backbone. We report AUPRC, AUROC,

Table 12: Comparison with task-specific EEG methods across five downstream datasets. Values are B-Acc (mean ± standard deviation, %).

Method	TUEV	TUAB	SEED-V	BCIC-2a	HMC
EEGNet	38.76±1.43	76.42±0.36	29.61±1.02	44.82±0.94	65.34±1.22
EEGConformer	40.74±1.64	77.58±0.49	35.37±1.12	46.96±1.06	71.49±0.86
SPaRCNet	41.61±2.62	78.96±0.18	29.49±0.78	46.35±1.17	47.56±11.09
ContraWR	43.84±3.49	77.46±0.41	35.46±1.05	46.78±1.25	42.42±5.41
CNN-Transformer	40.87±1.61	77.77±0.22	36.78±0.78	46.00±1.08	65.73±1.41
FFCL	39.79±1.04	78.48±0.38	36.41±0.92	44.70±1.43	44.27±7.02
ST-Transformer	39.84±2.28	79.66±0.23	30.52±0.72	45.75±1.45	25.59±1.41
CSPNet	39.50±7.40	77.90±0.50	–	–	–
TS-SEFFNet	48.60±0.90	80.00±0.30	–	–	–
MSCFormer	42.60±1.20	79.70±0.50	–	–	–
CTNet	27.10±2.10	79.00±1.50	–	–	–
HRAGP (LaBraM)	67.67±1.90	83.01±0.25	42.66±0.55	37.63±0.86	75.36±0.59
HRAGP (EEGPT)	64.37±1.65	80.99±0.40	39.76±0.85	46.16±0.96	74.59±0.30
HRAGP (CBraMod)	69.36±1.43	82.22±0.57	31.27±1.44	37.57±0.77	72.43±0.56
HRAGP (REVE)	78.46±1.88	82.17±0.16	43.86±1.16	54.69±1.65	74.27±0.63

and balanced accuracy (B-Acc). Each ablation group uses a fixed seed within the corresponding comparison, absolute values across different ablation tables are not directly comparable.

F.1 Extended Geometric Relation Conditions

We first provide the complete 2^3 combination of the three geometric conditions used in relation induction. The three conditions correspond to cone compatibility (C_1), directional preference (C_2), and radial ordering (C_3).

Table 13: Complete ablation of the three geometric relation conditions on TUAB with LaBraM-Base.

C_1	C_2	C_3	AUPRC	AUROC	B-Acc
✓	✓	✓	**0.9041**	**0.9050**	**0.8231**
✓	✓	×	0.9026	0.9044	0.8188
✓	×	✓	0.9023	0.9026	0.8197
✓	×	×	0.9023	0.9040	0.8194
×	✓	✓	0.8979	0.8991	0.8131
×	✓	×	0.9026	0.9027	0.8201
×	×	✓	0.9008	0.9013	0.8181
×	×	×	0.9031	0.9032	0.8184

F.2 Direction of Radial Ordering

The radial condition introduces an explicit ordering between source and target tokens. We therefore reverse the radial comparison to examine whether the orientation itself contributes to relation induction.

Table 14: Effect of radial-order direction on TUAB with LaBraM-Base.

Radial condition	**AUPRC**	**AUROC**	**B-Acc**
$r_j - r_i$	**0.9038**	**0.9048**	**0.8216**
$r_i - r_j$	0.8998	0.9016	0.8154

F.3 Graph Sparsity

We next examine the effect of graph sparsity. Different relation threshold configurations produce graphs with different fractions of retained directed edges.

Table 15: Effect of graph sparsity on TUAB with LaBraM-Base.

Graph sparsity	AUPRC	AUROC	B-Acc
0.0127	0.8977	0.8997	0.8165
0.0264	**0.9038**	**0.9048**	**0.8216**
0.0384	0.9021	0.9034	0.8196
0.0440	0.8970	0.8990	0.8142

F.4 Propagation Configuration

We further examine how additional propagation steps affect the directional graph readout. The default configuration uses two-hop propagation. We compare it with two types of deeper propagation extensions.

For the α-weighted variants, additional higher-order propagation features are added on top of the default two-hop representation with a weight α. In other words, the deeper-step contribution is scaled before being incorporated into the graph representation. For the decay-based variants, propagation features are progressively attenuated along the propagation path, such that information from deeper steps receives smaller weights.

Table 16: Comparison of propagation configurations on TUAB with LaBraM-Base. "Steps" denotes the additional propagation depth. α controls the contribution of higher-order propagation features, while "decay" progressively reduces the weight of features along deeper propagation paths.

Configuration	AUPRC	AUROC	B-Acc
Two-hop	0.9009	0.9009	**0.8198**
3 steps, $\alpha = 0.2$	0.9012	0.9013	0.8181
5 steps, $\alpha = 0.2$	**0.9013**	**0.9018**	0.8156
5 steps, $\alpha = 0.35$	0.8966	0.8986	0.8145
3 steps, decay $= 0.5$	0.8990	0.8969	0.8148
4 steps, decay $= 0.65$	0.8925	0.8096	0.8060

G Limitations and Future Work

HRAGP has two main limitations. First, its geometric modeling requires an analysis of the pretrained token space before selecting an appropriate manifold and relation formulation, which introduces additional modeling considerations beyond standard downstream readouts. Second, the geometric thresholds used for graph induction are shared design choices and may not be optimal for every individual downstream task.

Future work will investigate whether similar relation-aware readout strategies can benefit from other geometric manifolds, and more broadly explore how different geometric structures correspond to the semantic organization of pretrained EEG representations.