PSEUDO-HYPERBOLIC NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Standard hyperbolic neural networks typically rely on a single timelike dimension, providing a strong inductive bias for hierarchical data but limiting their flexibility in modeling heterogeneous data with mixed structural patterns. In this work, we investigate representation learning on pseudo-hyperbolic manifolds with multiple timelike dimensions. Our main contribution is the proposal of Pseudo-Hyperbolic Multinomial Logistic Regression (QMLR), which, to our knowledge, is the first MLR classifier formulated directly on pseudo-hyperbolic manifolds. Specifically, QMLR derives an exact spacelike point-to-hyperplane decision function and constructs a timelike surrogate for cases where the corresponding global minimum-distance formulation is not well-defined. Class-dependent offsets are encoded through generalized Lorentz boosts, while a learnable gate adaptively combines the two causal branches. To enhance the representations supplied to QMLR, we further develop a pseudo-hyperbolic attention module that integrates geometry-aware aggregation, radial projection, and learnable residual interpolation while preserving the manifold constraint. Extensive experiments on graph node classification, electroencephalography (EEG) decoding, image classification, and genomic sequence learning demonstrate that our proposed framework outperforms the compared baselines in most settings, supporting pseudo-hyperbolic geometry as a flexible alternative for heterogeneous representation learning.

1 INTRODUCTION

Geometric deep learning extends representation learning beyond Euclidean spaces by incorporating the intrinsic data structure (Bronstein et al., 2021). Hyperbolic geometry has been adopted for hierarchical representation learning through the Poincaré and Lorentz models (Nickel & Kiela, 2017; 2018; Ganea et al., 2018; Chami et al., 2019). However, Lorentzian models use a single timelike dimension, which may be restrictive when heterogeneous structural patterns coexist. Pseudo-Riemannian spaces provide a more flexible alternative by allowing indefinite metrics with multiple timelike dimensions (Law & Stam, 2020; Sim et al., 2021; Xiong et al., 2022; Le & Ta, 2025).

However, native classification on pseudo-hyperbolic manifolds remains underexplored. Existing hyperbolic classifiers extend Euclidean MLR through geometric decision functions in Poincaré and Lorentz spaces (Ganea et al., 2018; Shimizu et al., 2021; Bdeir et al., 2024; Nguyen et al., 2025; Chen et al., 2026a; Shi et al., 2026). These formulations assume Riemannian hyperbolic geometry, whereas pseudo-hyperboloids admit both spacelike and timelike causal directions, for which a single distance-based classifier does not apply uniformly.

We address the gap by introducing Pseudo-Hyperbolic Multinomial Logistic Regression (QMLR), an MLR classifier defined directly on pseudo-hyperboloid. For spacelike sections, we formulate the classifier from signed point-to-hyperplane geodesic distances and use generalized Lorentz boost to encode class dependent offsets while preserving the indefinite metric. An analogous global minimum-distance formulation is not well-defined for timelike normals; therefore, we introduce a timelike surrogate score. A learnable gate combines the two branches, allowing the decision function to accommodate geometric characteristics. To refine representations before classification, we further introduce a pseudo-hyperbolic attention module that combines geometry-aware aggregation in the ambient pseudo-Euclidean space with radial projection and residual interpolation, thereby preserving the pseudo-hyperboloid constraint throughout the refinement process.

Our contributions are summarized as follows:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

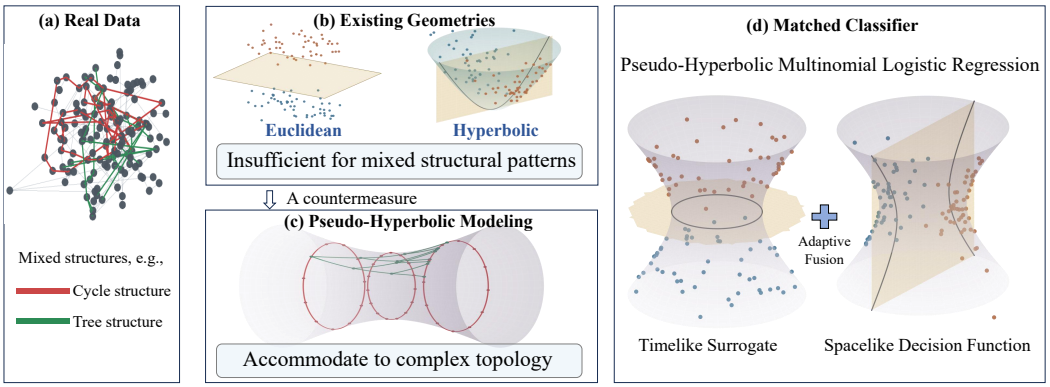


Figure 1: Motivation for the proposed pseudo-hyperbolic framework. (a) Real-world data may exhibit heterogeneous structures, including tree-like hierarchies and cycle-rich patterns. (b) Euclidean geometry is flat, whereas conventional hyperbolic models are associated with a single ambient timelike direction and primarily favor hierarchical structures. (c) Pseudo-hyperbolic geometry introduces multiple timelike directions, providing greater flexibility for modeling mixed structural patterns. (d) QMLR performs classification directly on the pseudo-hyperboloid by adaptively fusing an exact spacelike point-to-hyperplane decision score with a timelike surrogate.

- **Native pseudo-hyperbolic classification.** We introduce QMLR, which, to our knowledge, is the first MLR classifier defined directly on pseudo-hyperbolic manifolds. It combines an exact spacelike signed-distance score with a timelike surrogate through adaptive gating, while encoding class offsets via generalized Lorentz boosts.
- **Geometry-preserving feature refinement**. We develop a pseudo-hyperbolic attention module that integrates geometry-aware ambient aggregation, radial projection, and learnable residual interpolation, enabling adaptive feature refinement without tangent-space approximations.
- **Extensive experimental validation**. Experiments across graph node classification, EEG decoding, image classification, and genomic analysis demonstrate the effectiveness of QMLR and pseudo-hyperbolic attention, with competitive or superior performance over Euclidean and conventional hyperbolic baselines.

2 Related Work

Hyperbolic representation learning. Hyperbolic geometry has been widely used for hierarchical representation learning, from early Poincaré and Lorentz embeddings (Nickel & Kiela, 2017; 2018) to hyperbolic neural operations and graph networks (Ganea et al., 2018; Chami et al., 2019; Liu et al., 2019). Subsequent studies reduced reliance on tangent-space approximations through fully hyperbolic architectures (Chen et al., 2022; Bdeir et al., 2024; Shi et al., 2026), while alternative parameterizations such as Proper Velocity space have also been explored (Chen et al., 2026b). Alongside these advances, attention mechanisms have also been adapted to the underlying geometry, with formulations developed for the hyperboloid, Klein, and Poincaré models (Shimizu et al., 2021; Gulcehre et al., 2019). These developments establish a broad foundation for native hyperbolic representation learning.

Hyperbolic classification. Euclidean MLR admits a geometric interpretation through signed distances to class-specific hyperplanes (Lebanon & Lafferty, 2004). This interpretation has been extended to Poincaré and Lorentz models (Ganea et al., 2018; Shimizu et al., 2021; Bdeir et al., 2024). More recent classifiers construct decision scores using symmetric-space distances, Busemann functions, and intrinsic Lorentz hyperplanes (Nguyen et al., 2025; Chen et al., 2026a; Shi et al., 2026). These methods are designed for Riemannian hyperbolic manifolds. On a pseudo-hyperboloid with an indefinite induced metric, however, decision normals may have different causal types, and a single point-to-hyperplane construction does not apply uniformly to both spacelike and timelike cases.

Pseudo-Riemannian representation learning. Pseudo-Riemannian models relax the positive-definiteness requirement through indefinite metric signatures. Law & Stam (2020) introduced ultrahyperbolic representation learning with constant nonzero curvature, while Sim et al. (2021) studied directed graph embeddings in pseudo-Riemannian spaces. Subsequent work developed pseudo-Riemannian graph convolutional networks and Transformers for learning representations of graphs with mixed topological structures (Xiong et al., 2022; Le & Ta, 2025). These methods primarily focus on representation construction and feature propagation, leaving geometry-aware multiclass classification on pseudo-hyperboloids largely unexplored. In this work, our proposed QMLR explicitly models spacelike and timelike structures and combines them via adaptive fusion.

3 PRELIMINARIES

A pseudo-Riemannian manifold $(\mathcal{M}, g)$ is a smooth manifold $\mathcal{M}$ endowed with a symmetric, non-degenerate metric tensor g of constant index (O'neill, 1983). At each $x \in \mathcal{M}$, the metric induces an indefinite inner product $g_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$ on the tangent space $T_x\mathcal{M}$. Non-degeneracy requires $g_x(v, w) = 0$ for all $w \in T_x\mathcal{M}$ only when $v = 0$. By Sylvester's law of inertia, g_x admits a basis with signature (p, q), represented by $\eta_{p,q} = \mathrm{diag}(-I_p, I_q)$. Accordingly, each tangent space is locally modeled by the pseudo-Euclidean space $\mathbb{R}^{p,q}$, whose inner product is

$$\langle u, v\rangle_{p,q} = -\sum_{i=1}^{p} u_i v_i + \sum_{j=p+1}^{p+q} u_j v_j. \tag{1}$$

A nonzero vector v is timelike, spacelike, or null according to $\langle v, v\rangle_{p,q} < 0, > 0$, or $= 0$. The cases $p = 0$ and $p = 1$ recover Riemannian and Lorentzian geometry.

Pseudo-hyperboloid. In the ambient pseudo-Euclidean space $\mathbb{R}^{t+1,s}$, corresponding to $p = t + 1$ and $q = s$ above. For $x, y \in \mathbb{R}^{t+1,s}$, we define pseudo-hyperbolic inner product as

$$\langle x, y\rangle_{\mathcal{Q}} = -\sum_{i=0}^{t} x_i y_i + \sum_{j=t+1}^{t+s} x_j y_j, \tag{2}$$

which has ambient signature $(t + 1, s)$. For $\beta < 0$, the pseudo-hyperboloid is the smooth level set

$$\mathcal{Q}_\beta^{t,s} = \left\{ x \in \mathbb{R}^{t+1,s} \;\middle|\; \langle x, x\rangle_{\mathcal{Q}} = \beta \right\}. \tag{3}$$

Its tangent space satisfies $T_x\mathcal{Q}_\beta^{t,s} = \{v \in \mathbb{R}^{t+1,s} \mid \langle x, v\rangle_{\mathcal{Q}} = 0\}$. Since $\beta < 0$, the radial normal x is timelike; restricting the ambient metric to its pseudo-orthogonal complement therefore removes one timelike direction.

4 PROPOSED METHOD

4.1 PSEUDO-HYPERBOLIC MODELING

The pipeline of our proposed framework is illustrated in Fig. 2. Given an input x, a feature decoder f_Θ produces an intermediate representation $\tilde{\mathbf{z}} = f_\Theta(x) = (\mathbf{z}_t, \mathbf{z}_s) \in \mathbb{R}^t \times \mathbb{R}^s$. Following the coordinate ordering of the ambient space $\mathbb{R}^{t+1,s} = \mathbb{R} \times \mathbb{R}^t \times \mathbb{R}^s$, we choose the canonical base point $\mathbf{o} = (\sqrt{-\beta}, 0, \ldots, 0)^\top \in \mathcal{Q}_\beta^{t,s}$ as the origin and construct $\mathbf{z} = (0, \tilde{\mathbf{z}})^\top$. Since $\langle \mathbf{o}, \mathbf{z}\rangle_{\mathcal{Q}} = 0$, we have $\mathbf{z} \in \mathcal{T}_{\mathbf{o}}\mathcal{Q}_\beta^{t,s}$. The initial pseudo-hyperbolic representation is then defined as $\mathbf{h}^{(0)} = \mathrm{Exp}_{\mathbf{o}}^{\mathcal{Q}}(\mathbf{z})$.

Because the pseudo-Riemannian metric is indefinite, the exponential map takes different forms depending on the causal type of $\mathbf{z}$. Specifically, $\langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} > 0$, $\langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} < 0$, and $\langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} = 0$ correspond to spacelike, timelike, and lightlike (null) tangent directions, respectively. For a non-null tangent vector $\mathbf{z}$, let $\rho = \sqrt{|\langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}}|}/\sqrt{-\beta}, \beta < 0, \|\mathbf{z}\|_{\mathcal{Q}} := \sqrt{|\langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}}|}$, the exponential map is then given by:

$$\mathrm{Exp}_{\mathbf{o}}^{\mathcal{Q}}(\mathbf{z}) = \begin{cases} \cosh(\rho)\,\mathbf{o} + \sqrt{-\beta}\,\sinh(\rho)\,\dfrac{\mathbf{z}}{\|\mathbf{z}\|_{\mathcal{Q}}}, & \langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} > 0, \\ \cos(\rho)\,\mathbf{o} + \sqrt{-\beta}\,\sin(\rho)\,\dfrac{\mathbf{z}}{\|\mathbf{z}\|_{\mathcal{Q}}}, & \langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} < 0, \\ \mathbf{o} + \mathbf{z}, & \langle \mathbf{z}, \mathbf{z}\rangle_{\mathcal{Q}} = 0. \end{cases} \tag{4}$$

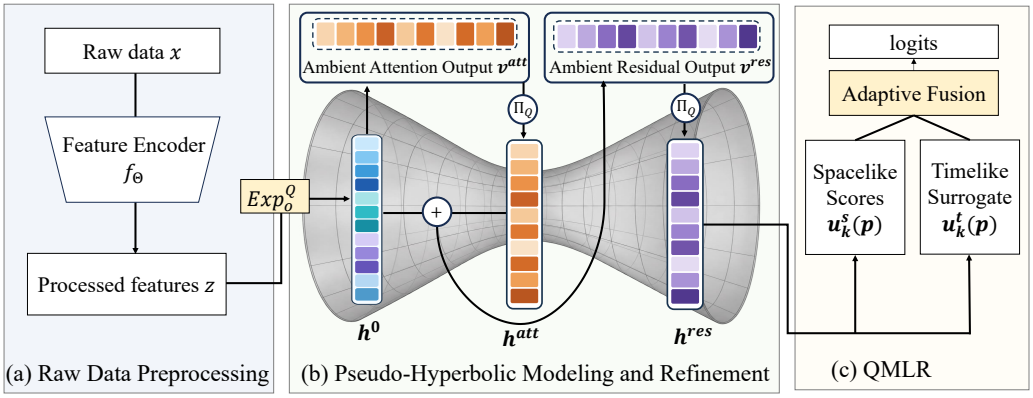


Figure 2: The pipeline of the proposed pseudo-hyperbolic framework. (a) Raw input data are processed by a feature encoder to obtain intermediate features, which are mapped to the pseudo-hyperboloid through the exponential map. (b) The resulting pseudo-hyperbolic representations are refined by geometry-based attention and residual interpolation. Attention aggregation and residual interpolation are performed in the ambient space, followed by radial projection back to the pseudo-hyperboloid. (c) The refined manifold representations are directly classified by QMLR, which adaptively fuses spacelike scores and a timelike surrogate to produce the final logits.

For a collection of intermediate features $\{\mathbf{z}_i\}_{i=1}^N$, the resulting pseudo-hyperbolic representations are denoted by $\mathbf{H}^{(0)} = \{\mathbf{h}_i^{(0)}\}_{i=1}^N$, where $\mathbf{h}_i^{(0)} = \mathrm{Exp}_{\mathbf{o}}^{\mathcal{Q}}(\mathbf{z}_i)$.

4.2 Pseudo-Hyperbolic Attention

Although the preceding pseudo-hyperbolic operations preserve the geometry of individual representations, they do not explicitly capture contextual interactions among features. We therefore introduce a geometry-aware attention mechanism that adaptively aggregates information from related features to refine each representation while preserving the pseudo-hyperbolic constraint. Let $\{\mathbf{h}_i^{(0)}\}_{i=1}^N \subset \mathcal{Q}_\beta^{t,s}$ be a causally coherent set satisfying $\langle \mathbf{h}_j^{(0)}, \mathbf{h}_k^{(0)} \rangle_{\mathcal{Q}} \le \beta < 0$ for all j, k. This admissibility condition ensures that positive ambient combinations of these representations remain timelike. Within this region, we define the pairwise geometric affinity using the pseudo-Riemannian inner product as $s_{ij} = \langle \mathbf{h}_i^{(0)}, \mathbf{h}_j^{(0)} \rangle_{\mathcal{Q}} / (-\beta) + 1$. The corresponding attention weight can be formulated as:

$$\alpha_{ij} = \frac{\exp(\tau s_{ij})}{\sum_{k=1}^N \exp(\tau s_{ik})}, \qquad \tau > 0. \tag{5}$$

The admissibility condition gives $s_{ij} \le 0$, with $s_{ii} = 0$. Derivation is provided in App. B.6. Therefore, representations with larger affinity scores receive greater attention weights. The weighted combination induced by α_{ij}, *i.e.*, $\mathbf{v}_i^{\mathrm{att}} = \sum_{j=1}^N \alpha_{ij} \mathbf{h}_j^{(0)}$, is computed in the ambient space and may leave the manifold. Since $\alpha_{ij} \ge 0$ and $\sum_j \alpha_{ij} = 1$, the admissibility condition gives $\langle \mathbf{v}_i^{\mathrm{att}}, \mathbf{v}_i^{\mathrm{att}} \rangle_{\mathcal{Q}} = \sum_{j=1}^N \sum_{k=1}^N \alpha_{ij}\alpha_{ik} \langle \mathbf{h}_j^{(0)}, \mathbf{h}_k^{(0)} \rangle_{\mathcal{Q}} \le \beta \sum_{j=1}^N \sum_{k=1}^N \alpha_{ij}\alpha_{ik} = \beta < 0$. Hence, $\mathbf{v}_i^{\mathrm{att}}$ remains timelike, and we utilize the radial projection $\Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{att}})$, which is similar to the ambient normalization used for aggregation in Lorentz model (Law et al., 2019), to map the aggregated data back onto the manifold:

$$\Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{att}}) = \sqrt{\frac{-\beta}{|\langle \mathbf{v}_i^{\mathrm{att}}, \mathbf{v}_i^{\mathrm{att}} \rangle_{\mathcal{Q}}|}}\, \mathbf{v}_i^{\mathrm{att}} \tag{6}$$

It follows that $\langle \Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{att}}), \Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{att}}) \rangle_{\mathcal{Q}} = \beta$. The attention output is given by $\mathbf{h}_i^{\mathrm{att}} = \Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{att}})$. Although attention aggregation incorporates contextual information, it may weaken the original representation. We therefore use a residual interpolation to balance the original and attention-enhanced features. Let $\theta = \sigma(\lambda) \in (0, 1)$, where λ is a learnable scalar, the residual representation and the subsequent radial projection can be written as:

$$\mathbf{v}_i^{\mathrm{res}} = \theta \mathbf{h}_i^{(0)} + (1 - \theta) \mathbf{h}_i^{\mathrm{att}}, \quad \mathbf{h}_i^{\mathrm{res}} = \Pi_{\mathcal{Q}}(\mathbf{v}_i^{\mathrm{res}}). \tag{7}$$

The interpolation remains timelike for $\theta \in [0,1]$ because $\mathbf{h}_i^{(0)}$ and $\mathbf{h}_i^{\mathrm{att}}$ belong to the same causal component. Thereby, $\Pi_{\mathcal{Q}}$ is well-defined and maps $\mathbf{v}_i^{\mathrm{res}}$ back to $\mathcal{Q}_\beta^{t,s}$. A proof is given in App. B.6.

4.3 Pseudo-Hyperbolic Multinomial Logistic Regression

4.3.1 Geometric Formulation

The Euclidean MLR $\mathrm{Softmax}(Ax+b)$ can be reformulated through signed point-to-hyperplane distances. For class k, let $a_k \in \mathbb{R}^n$ and $b_k \in \mathbb{R}$ denote the weight and bias. Define $w_k = a_k/\|a_k\|_2$, $\bar{b}_k = b_k/\|a_k\|_2$ and $H_k^E = \{x \in \mathbb{R}^n \mid \langle w_k, x\rangle + \bar{b}_k = 0\}$. The class score can be written as

$$u_k^E(x) = \|a_k\|_2 \,\mathrm{sign}\big(\langle w_k, x\rangle + \bar{b}_k\big)\, d_E(x, H_k^E), \tag{8}$$

We extend this formulation to the pseudo-hyperboloid $\mathcal{Q}_\beta^{t,s}$. For a nonzero ambient normal a, the corresponding decision section is

$$H_a^{\mathcal{Q}} = \Big\{ x \in \mathcal{Q}_\beta^{t,s} \mid \langle x, a\rangle_{\mathcal{Q}} = 0 \Big\}. \tag{9}$$

To incorporate the class offset, we parameterize the normal through a generalized Lorentz boost. With $\phi_k = -\bar{b}_k/\sqrt{-\beta}$, let $v_k \in \mathbb{R}^{t+1}$ and $u_k \in \mathbb{R}^s$ denote unit temporal and spatial directions.Details are provided in App. C.3, the generalized Lorentz boost is given by

$$B_k(\phi_k) = \begin{pmatrix} I_{t+1} + (\cosh\phi_k - 1) v_k v_k^\top & \sinh\phi_k\, v_k u_k^\top \\ \sinh\phi_k\, u_k v_k^\top & I_s + (\cosh\phi_k - 1) u_k u_k^\top \end{pmatrix}. \tag{10}$$

The transformation satisfies $B_k^\top G B_k = G$ and is therefore pseudo-orthogonal. Starting from the canonical unit spacelike normal $a_k^0 = (0^\top, u_k^\top)^\top$, we set $\tilde{a}_k^s = B_k(\phi_k) a_k^0$. Hence $\langle \tilde{a}_k^s, \tilde{a}_k^s\rangle_{\mathcal{Q}} = 1$, and the spacelike decision section is $H_k^{\mathcal{Q},s} = \{q \in \mathcal{Q}_\beta^{t,s} \mid \langle q, \tilde{a}_k^s\rangle_{\mathcal{Q}} = 0\}$.

4.3.2 Spacelike Decision Function

We define the point-to-section distance within an admissible spacelike cluster:

$$\begin{aligned} &d_{\mathcal{Q}}^s(p, H_k^{\mathcal{Q},s}) := \inf_{q \in \mathcal{A}_k^s(p)} d_{\mathcal{Q}}^s(p,q), \\ &q^\star \in \mathcal{A}_k^s(p) \subseteq H_k^{\mathcal{Q},s}, \qquad \langle u, v\rangle_{\mathcal{Q}} < \beta \quad (\forall u \neq v \in \mathcal{A}_k^s(p)). \end{aligned} \tag{11}$$

Where $q^\star$ is the normal foot defined in App. C.4, and $d_{\mathcal{Q}}^s(p,q)$ denotes the spacelike geodesic length, with $d_{\mathcal{Q}}^s(p,p) = 0$.

Theorem 1 (Spacelike Point-to-Hyperplane Distance) *For a unit spacelike normal $\tilde{a}_k^s$, under the cluster conditions in Eq. (11), we can obtain the following*

$$d_{\mathcal{Q}}^s\Big(p, H_k^{\mathcal{Q},s}\Big) = \sqrt{-\beta}\, \left|\mathrm{asinh}\left(\frac{\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}}}{\sqrt{-\beta}}\right)\right|. \tag{12}$$

Theorem 2 (Spacelike QMLR) *For $p = (p_t, p_s) \in \mathcal{Q}_\beta^{t,s}$, the spacelike QMLR score can be obtained from the signed-distance shown below:*

$$u_k^s(p) = \|a_k\|_2\, \mathrm{sign}(\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}})\, d_{\mathcal{Q}}^s\Big(p, H_k^{\mathcal{Q},s}\Big) \quad = \|a_k\|_2 \sqrt{-\beta}\, \mathrm{asinh}\left(\frac{\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}}}{\sqrt{-\beta}}\right). \tag{13}$$

We parameterize the unit spacelike normal by $\hat{a}_{k,t} = a_{k,t}/\|a_{k,t}\|_2$, $\hat{a}_{k,s} = a_{k,s}/\|a_{k,s}\|_2$, and $\tilde{a}_k^s = \big(\sinh(\phi_k)\hat{a}_{k,t}^\top, \cosh(\phi_k)\hat{a}_{k,s}^\top\big)^\top$, where $\phi_k = -\bar{b}_k/\sqrt{-\beta}$. Then

$$u_k^s(p) = \|a_k\|_2 \sqrt{-\beta}\, \mathrm{asinh}\left[\frac{1}{\sqrt{-\beta}}\left(\cosh(\phi_k)\frac{\langle p_s, a_{k,s}\rangle}{\|a_{k,s}\|_2} - \sinh(\phi_k)\frac{\langle p_t, a_{k,t}\rangle}{\|a_{k,t}\|_2}\right)\right]. \tag{14}$$

4.3.3 TIMELIKE DECISION FUNCTION

The corresponding global minimum-distance construction in timelike space is not well-defined. We therefore use a surrogate score with the same curvature-aware mapping. Writing $p = (p_t, p_s)$, with $p_t \in \mathbb{R}^{t+1}$ and $p_s \in \mathbb{R}^s$, define

$$z_k^t(p) = \cosh(\phi_k)\frac{\langle p_t, a_{k,t}\rangle}{\|a_{k,t}\|_2} - \sinh(\phi_k)\frac{\langle p_s, a_{k,s}\rangle}{\|a_{k,s}\|_2}.$$

Unlike u_k^s, the score u_k^t is a surrogate class score which can be written as

$$u_k^t(p) = \|a_k\|_2\sqrt{-\beta}\ \operatorname{asinh}\left(\frac{z_k^t(p)}{\sqrt{-\beta}}\right). \tag{15}$$

4.3.4 ADAPTIVE FUSION

Having obtained the class scores from the spacelike and timelike branches, we combine their respective contributions using a learnable gate $g = \sigma(\gamma)$:

$$u_k(p) = (1-g)u_k^s(p) + g\,u_k^t(p). \tag{16}$$

The resulting fused scores are then converted into class probabilities using the softmax function:

$$p(y = k \mid p) = \frac{\exp(u_k(p))}{\sum_{c=1}^{C}\exp(u_c(p))}.$$

For $\beta \to -\infty$, the spacelike branch recovers the Euclidean MLR score, as derived in App. C.7.

5 EXPERIMENTS

To evaluate our model under appropriate and controlled settings, we tailor the experimental configuration to the established architectures and baselines of each task. For EEG decoding, we evaluate the complete Pseudo-Hyperbolic Neural Network (Q-Net) following the common hybrid paradigm of existing manifold-based EEG networks: a Euclidean EEG encoder first extracts features, which are then mapped to $\mathcal{Q}_\beta^{t,s}$ and refined by pseudo-hyperbolic attention and residual interpolation. For node classification, where pseudo-hyperbolic graph backbones are already available, we adopt QGCN2 (Le & Ta, 2025) as the backbone and replace only its linear classifier with QMLR, thereby isolating the contribution of native pseudo-hyperbolic classification. For image and genomic tasks, following the controlled comparison protocol of (Chen et al., 2026a), we keep the backbone fixed and compare QMLR with MLR variants defined in other hyperbolic geometries. Unless otherwise stated, results are reported as mean $\pm$ standard deviation over 10 independent runs.

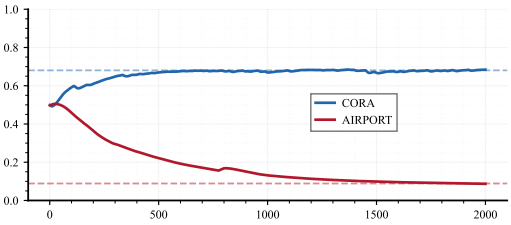
5.1 GRAPH NODE CLASSIFICATION

Results. We evaluate QMLR on Airport, CiteSeer, and Arxiv using QGCN2 (Le & Ta, 2025) as the shared pseudo-hyperbolic backbone. To isolate the effect of the classifier, we replace only its original linear head with QMLR while keeping other settings unchanged. As shown in Table 1, QMLR improves the mean F1 score by 2.09%, 1.00%, and 0.33% on Airport, CiteSeer, and Arxiv. These results show that QMLR serves as an effective classifier for pseudo-hyperbolic representations.

Ablation. To isolate the behavior of the two QMLR branches, we follow the best pseudo-hyperbolic signatures reported for the QGCN backbone on Cora and Airport (Xiong et al., 2022), then keep the backbone geometry fixed and compare the spacelike-only, timelike-only, and adaptively fused QMLR variants. We analyze two branches of QMLR on Cora and Airport, which exhibit different levels of Gromov hyperbolicity. As shown in Fig. 3, Cora favors the timelike branch, whereas Airport favors the spacelike branch. On adaptive-fused QMLR the learned gate converges to 0.6804 on Cora and 0.0888 on Airport, assigning greater weight to the corresponding preferred branch. Adaptive QMLR also achieves the best performance on both datasets, supporting the use of adaptive fusion.

Table 1: Performance comparison of different models on node classification tasks. Results are reported as F1-scores (%). MLR denotes a linear classifier.

Models	AIRPORT	CITESEER	ARXIV
GCN (Kipf & Welling, 2017)	81.40 ± 0.60	70.73 ± 0.61	57.39 ± 0.54
HGCN (Chami et al., 2019)	81.40 ± 0.60	65.60 ± 0.64	62.16 ± 0.36
κ-GCN (Yang et al., 2023)	80.60 ± 0.39	67.27 ± 0.71	58.36 ± 0.18
NodeFormer (Wu et al., 2022)	88.87 ± 0.36	70.03 ± 0.31	53.37 ± 0.34
SGFormer (Wu et al., 2023)	93.32 ± 1.09	71.93 ± 0.59	61.38 ± 0.25
Hypformer (Yang et al., 2024)	89.19 ± 0.50	67.77 ± 0.41	64.62 ± 0.22
QGCN2 + MLR (Le & Ta, 2025)	91.66 ± 0.45	71.63 ± 0.29	64.93 ± 0.32
QGCN2 + QMLR	**93.75** ± 0.43	**72.63** ± 0.70	**65.26** ± 0.21



(a) Gate convergence

Method	Cora $\mathcal{Q}^{14,2}$	Airport $\mathcal{Q}^{1,15}$
Spacelike-Only	67.16±6.54	91.73±0.88
Timelike-Only	78.34±1.03	89.86±5.28
QMLR	**81.72**±0.42	**93.75**±0.43

(b) Branch comparison

Figure 3: Adaptive behavior of QMLR on Cora and Airport. (a) Convergence of the learned gate coefficient g. (b) Comparison of the spacelike, timelike, and adaptively fused QMLR branches.

Table 2: Cross-session EEG decoding performance. MI is evaluated by accuracy (%), while SSVEP and ERN are evaluated by area under the ROC curve (AUC, %).

Geometry	Methods	MI	SSVEP	ERN
Euclidean	EEGNet (Lawhern et al., 2018)	61.84 ± 6.39	53.72 ± 7.23	74.28 ± 2.47
	ShallowConvNet (Schirrmeister et al., 2017)	57.43 ± 6.25	56.93 ± 6.97	71.86 ± 2.64
	SCCNet (Wei et al., 2019)	71.95 ± 5.05	62.11 ± 7.70	70.93 ± 2.31
	FBCNet (Mane et al., 2021)	56.52 ± 3.07	53.09 ± 5.67	60.47 ± 3.06
	TCNet-Fusion (Musallam et al., 2021)	71.45 ± 4.45	45.00 ± 6.45	70.46 ± 2.94
	MBEEGSE (Altuwaijri et al., 2022)	64.58 ± 6.07	56.45 ± 7.27	75.46 ± 2.34
Manifold	SPDNet (Huang & Van Gool, 2017)	72.93 ± 4.33	62.30 ± 3.12	72.05 ± 4.43
	SPDNetBN (Brooks et al., 2019)	73.02 ± 3.67	62.76 ± 3.01	72.34 ± 3.46
	MAtt (Pan et al., 2022)	74.71 ± 5.01	65.19 ± 3.14	75.68 ± 2.23
	GDLNet (Wang et al., 2024)	69.32 ± 2.89	65.52 ± 2.86	78.23 ± 2.52
	HEEGNet (Li et al., 2026)	68.73 ± 1.38	64.22 ± 1.50	75.16 ± 0.84
	Q-Net (Ours)	**74.86 ± 1.58**	**67.91 ± 0.95**	**82.62 ± 0.54**

5.2 EEG Cross-Session Decoding

Setup. We evaluate Q-Net on three EEG benchmarks covering Motor Imagery (MI) (Brunner et al., 2008), Steady-State Visual Evoked Potentials (SSVEP) (Nikolopoulos, 2021), and Error-Related Negativity (ERN) (Margaux et al., 2012). Following the hybrid paradigm, a Euclidean backbone is used as the feature encoder. Its outputs are mapped to $\mathcal{Q}_\beta^{t,s}$, refined through pseudo-hyperbolic attention and residual interpolation, and classified by QMLR. We report accuracy for MI and AUC for SSVEP and ERN. More details are provided in App. D.1.

Main results. As shown in Table 2, Q-Net remains competitive across all three benchmarks, with modest gains on MI and more pronounced improvements on SSVEP and ERN. These results show the effectiveness of the proposed pseudo-hyperbolic framework across different EEG paradigms.

Ablation studies. Table 3 evaluates the contributions of representation geometry, classifier design, and pseudo-hyperbolic attention. Moving from the linear classifier to Lorentz and pseudo-hyperbolic

Table 3: Ablation study on the key components of Q-Net.

Configuration	MI	SSVEP	ERN
Linear Classifier	58.43 ± 2.03	58.55 ± 1.68	81.48 ± 0.46
Lorentz + LMLR	68.89 ± 1.66	63.56 ± 0.92	81.67 ± 0.64
Pseudo-hyp. + Att. + Linear	70.69 ± 1.53	66.39 ± 0.63	82.46 ± 0.47
Pseudo-hyp. + QMLR	74.75 ± 1.69	61.95 ± 0.95	78.66 ± 0.77
Pseudo-hyp. + Att. + QMLR	**74.86 ± 1.58**	**67.91 ± 0.95**	**82.62 ± 0.54**

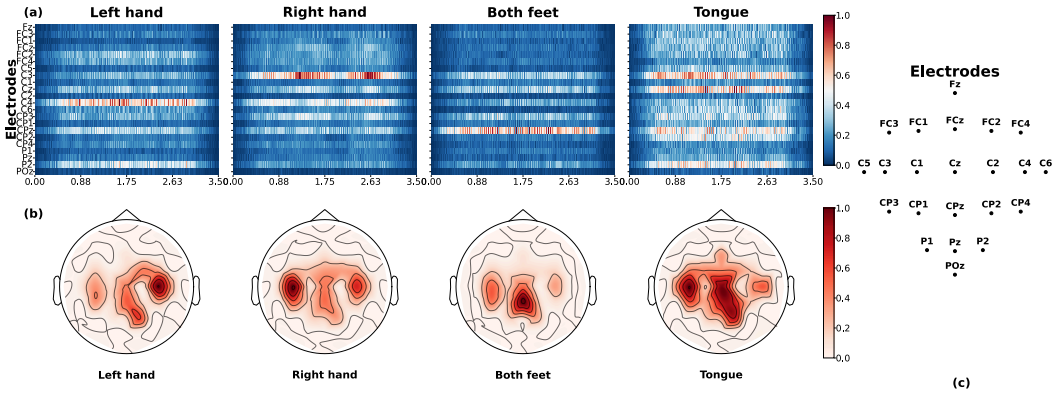


Figure 4: Gradient-based spatiotemporal saliency on the MI dataset. (a) Temporal saliency across EEG channels for four motor imagery classes. (b) Corresponding scalp topographies. (c) Electrode configuration used for visualization.

representations yields clear gains, while QMLR further improves the pseudo-hyperbolic model. Attention provides additional benefits on SSVEP and ERN, and the full model achieves the best performance across all three benchmarks. Additional ablations are provided in App. D.1.4.

Model interpretation. Figure 4 shows class-dependent spatiotemporal saliency on MI. Hand imagery exhibits strong responses over central sensorimotor regions, while feet imagery shows more pronounced midline activity, broadly consistent with established motor-imagery EEG responses (Pfurtscheller & Andrew, 1999). Additional visualizations are provided in App. D.1.5.

5.3 Image Classification

Setup. Using a shared ResNet-18 (He et al., 2016) backbone, we replace the classifier and compare QMLR with Euclidean, Poincaré, and Lorentz MLR variants on CIFAR-10 and CIFAR-100 (Krizhevsky et al., 2009).

Results. As shown in Table 4, QMLR achieves the best mean accuracy on CIFAR-10 and CIFAR-100, reaching 95.43% and 78.14%. Although the improvements over the strongest baselines are modest, QMLR maintains comparable training time and classifier size. These results show that our proposed pseudo-hyperbolic classifier remains competitive with existing Euclidean and hyperbolic classifiers while introducing negligible additional computational and parameter overhead.

5.4 Genome Sequence Learning

Setup. Following the hyperbolic genome-learning setting (Khan et al., 2025), we evaluate QMLR on the Transposable Elements Benchmark (TEB) and Genome Understanding Evaluation (GUE) (Zhou et al., 2024) using a shared CNN backbone with settings fixed, only the geometry-specific classifier is replaced. All tasks are binary classification except Covid Variant, which contains nine classes. Further experimental details are provided in App. D.3.

Table 4: Classification accuracy, training time (seconds per epoch), and parameters on CIFAR-10 and CIFAR-100.

Space	Method	CIFAR-10			CIFAR-100		
		Acc (%)	Time	Params	Acc (%)	Time	Params
$\mathbb{R}^n$	MLR	95.14 ± 0.12	14.2	5.13K	77.72 ± 0.15	14.1	51.30K
$\mathbb{P}^n_K$	PMLR (Shimizu et al., 2021)	95.14 ± 0.11	16.3	5.14K	77.78 ± 0.09	33.8	51.40K
	PBMLR (Nguyen et al., 2025)	95.23 ± 0.08	17.1	10.24K	77.78 ± 0.15	37.9	102.40K
	BMLR-P (Chen et al., 2026a)	95.32 ± 0.14	14.5	5.14K	78.10 ± 0.35	14.5	51.40K
$\mathbb{L}^n_K$	LMLR (Bdeir et al., 2024)	94.98 ± 0.12	14.5	5.13K	78.03 ± 0.21	14.5	51.30K
	BMLR-L (Chen et al., 2026a)	95.25 ± 0.02	14.4	5.14K	78.07 ± 0.26	14.3	51.40K
$\mathbb{Q}^{t,s}_\beta$	**QMLR**	$\mathbf{95.43 \pm 0.05}$	15.0	5.15K	$\mathbf{78.14 \pm 0.32}$	14.9	51.50K

Table 5: Genomic Matthews Correlation Coefficient (MCC) performance of different MLR methods under the same CNN backbone.

Bench.	**Task**	**Dataset**	**MLR**	**PMLR**	**LMLR**	**BMLR-L**	**QMLR**
TEB	Retrotransposons	LTR Copia	54.73±1.45	55.84±3.14	58.26±1.80	55.73±2.79	**58.28±3.10**
		LINEs	70.63±1.24	74.72±2.58	**76.63±1.71**	74.79±2.31	76.02±1.50
		SINEs	85.15±1.64	86.61±1.99	87.36±1.77	86.95±1.97	**89.63±1.69**
	DNA transposons	CMC-EnSpm	72.18±0.32	71.27±2.00	72.69±0.71	71.25±1.84	**74.38±0.83**
		hAT-Ac	**87.45±0.90**	84.88±1.41	86.09±0.40	84.82±1.20	86.33±0.24
	Pseudogenes	processed	60.66±0.82	63.86±2.94	65.42±2.08	64.05±2.45	**67.04±1.22**
		unprocessed	51.94±2.69	70.21±2.92	70.12±2.91	69.43±1.12	**72.29±0.75**
GUE	Core Promoter Detection	tata	78.26±2.85	78.85±1.41	79.47±1.31	78.53±2.46	**79.93±1.13**
		notata	66.60±1.07	68.23±0.84	66.63±0.32	67.53±0.66	**69.39±0.31**
		all	66.47±0.74	68.37±0.78	66.17±1.12	67.34±0.67	**68.63±1.08**
	Promoter Detection	tata	78.58±3.39	79.41±0.22	80.54±1.32	81.32±1.84	**81.65±1.61**
		notata	90.81±0.51	**91.49±0.45**	89.12±0.65	91.42±0.60	90.86±0.46
		all	88.00±0.39	**89.63±0.26**	85.79±0.70	88.44±0.79	88.52±0.46
	Covid Variant	Covid	66.43±0.74	66.92±0.42	61.64±1.04	69.37±0.62	**69.43±0.71**

Results. As shown in Table 5, QMLR achieves the best mean MCC on several TEB and GUE tasks and remains competitive on others. It also frequently outperforms the Lorentz classifier on many of the evaluated tasks, although no single classifier dominates every benchmark. These results demonstrate the robustness of QMLR across heterogeneous genomic sequence tasks and support the potential benefit of pseudo-hyperbolic decision geometry.

6 Conclusion

This work introduced QMLR, which, to our knowledge, is the first MLR classifier defined directly on pseudo-hyperbolic manifolds. QMLR adaptively combines an exact spacelike decision function with a timelike surrogate, while the proposed pseudo-hyperbolic attention module refines representations through geometry-aware aggregation, radial projection, and residual interpolation. Experiments across graph, EEG, image, and genomic tasks show competitive or superior performance in most settings. Further analyses reveal dataset-dependent branch preferences and confirm the complementary effects of QMLR and pseudo-hyperbolic attention. Overall, this paper establishes a practical framework for representation learning and classification on manifolds with indefinite metric signatures.

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

Ghadir Ali Altuwaijri, Ghulam Muhammad, Hamdi Altaheri, and Mansour Alsulaiman. A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for eeg-based motor imagery signals classification. *Diagnostics*, 12(4):995, 2022.

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural networks for computer vision. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 47687–47711, 2024.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

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.

C Brunner, R Leeb, G Müller-Putz, A Schlögl, G Pfurtscheller, and BCI Competition. Graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16:1–6, 2008.

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

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5672–5686, 2022.

Ziheng Chen, Bernhard Schölkopf, and Nicu Sebe. Hyperbolic Busemann neural networks. In *CVPR*, pp. 42029–42038, 2026a.

Ziheng Chen, Zihan Su, Bernhard Schölkopf, and Nicu Sebe. Proper velocity neural networks. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 122721–122760, 2026b.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in neural information processing systems*, 31, 2018.

Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. Hyperbolic attention networks. In *International Conference on Learning Representations*, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

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.

Raiyan R. Khan, Philippe Chlenski, and Itsik Pe'er. Hyperbolic genome embeddings. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 73425–73454, 2025.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Marc Law and Jos Stam. Ultrahyperbolic representation learning. *Advances in neural information processing systems*, 33:1668–1678, 2020.

Marc Law, Renjie Liao, Jake Snell, and Richard Zemel. Lorentzian distance learning for hyperbolic representations. In *International conference on machine learning*, pp. 3672–3681. PMLR, 2019.

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.

Viet Quan Le and Viet Cuong Ta. Pseudo-riemannian graph transformer. 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. 15951–15983. Curran Associates, Inc., 2025. doi: 10.52202/085713-0539.

Guy Lebanon and John Lafferty. Hyperplane margin classifiers on the multinomial manifold. In *Proceedings of the twenty-first international conference on Machine learning*, pp. 66, 2004.

Shanglin Li, Shiwen Chu, Okan Koç, Yi Ding, Qibin Zhao, Motoaki Kawanabe, and Ziheng Chen. Heegnet: Hyperbolic embeddings for eeg. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 74317–74346, 2026.

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

Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A Prasad Vinod, Seong-Whan Lee, and Cuntai Guan. Fbcnet: A multi-view convolutional neural network for brain-computer interface. *arXiv preprint arXiv:2104.01233*, 2021.

Perrin Margaux, Maby Emmanuel, Daligault Sébastien, Bertrand Olivier, and Mattout Jérémie. Objective and subjective evaluation of online error correction during p300-based spelling. *Advances in Human-Computer Interaction*, 2012(1):578295, 2012.

Yazeed K Musallam, Nasser I AlFassam, Ghulam Muhammad, Syed Umar Amin, Mansour Alsulaiman, Wadood Abdul, Hamdi Altaheri, Mohamed A Bencherif, and Mohammed Algabri. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomedical Signal Processing and Control*, 69:102826, 2021.

Xuan Son Nguyen, Yang Yang, and Aymeric Histace. Neural networks on symmetric spaces of noncompact type. In *International Conference on Learning Representations*, volume 2025, pp. 96796–96832, 2025.

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.

Spiros Nikolopoulos. Mamem eeg ssvep dataset ii (256 channels, 11 subjects, 5 frequencies presented simultaneously). *Centre Res. Technol. Hellas, Inf. Technol. Inst., Thessaloniki, Greece, Tech. Rep*, pp. 5, 2021.

Barrett O'neill. *Semi-Riemannian geometry with applications to relativity*, volume 103. Academic press, 1983.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. *Advances in Neural Information Processing Systems*, 35:31116–31129, 2022.

Gert Pfurtscheller and Colin Andrew. Event-related changes of band power and coherence: methodology and interpretation. *Journal of clinical neurophysiology*, 16(6):512–519, 1999.

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.

Xianglong Shi, Ziheng Chen, Yunhan Jiang, and Nicu Sebe. Intrinsic lorentz neural network. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 107234–107256, 2026.

Ryohei Shimizu, Yusuke Mukuta, and Tatsuya Harada. Hyperbolic Neural Networks++. In *International Conference on Learning Representations*, 2021.

Aaron Sim, Maciej L Wiatrak, Angus Brayne, Páidí Creed, and Saee Paliwal. Directed graph embeddings in pseudo-riemannian manifolds. In *International Conference on Machine Learning*, pp. 9681–9690. PMLR, 2021.

Rui Wang, Chen Hu, Ziheng Chen, Xiao-Jun Wu, and Xiaoning Song. A grassmannian manifold self-attention network for signal classification. In *IJCAI*, pp. 5099–5107, 2024.

Chun-Shu Wei, Toshiaki Koike-Akino, and Ye Wang. Spatial component-wise convolutional network (sccnet) for motor-imagery eeg classification. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 328–331. IEEE, 2019.

Qitian Wu, Wentao Zhao, Zenan Li, David P Wipf, and Junchi Yan. Nodeformer: A scalable graph structure learning transformer for node classification. *Advances in neural information processing systems*, 35:27387–27401, 2022.

Qitian Wu, Wentao Zhao, Chenxiao Yang, Hengrui Zhang, Fan Nie, Haitian Jiang, Yatao Bian, and Junchi Yan. Sgformer: Simplifying and empowering transformers for large-graph representations. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 64753–64773. Curran Associates, Inc., 2023. doi: 10.52202/075280-2826.

Bo Xiong, Shichao Zhu, Nico Potyka, Shirui Pan, Chuan Zhou, and Steffen Staab. Pseudo-riemannian graph convolutional networks. *Advances in neural information processing systems*, 35:3488–3501, 2022.

Menglin Yang, Min Zhou, Lujia Pan, and Irwin King. κhgcn: Tree-likeness modeling via continuous and discrete curvature learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2965–2977, 2023.

Menglin Yang, Harshit Verma, Delvin Ce Zhang, Jiahong Liu, Irwin King, and Rex Ying. Hypformer: Exploring efficient transformer fully in hyperbolic space. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3770–3781, 2024.

Zhihan Zhou, Yanrong Ji, Weijian Li, Pratik Dutta, Ramana V Davuluri, and Han Liu. DNABERT-2: Efficient Foundation Model and Benchmark for Multi-Species Genomes. In *International Conference on Learning Representations*, 2024.

A NOTATION

Table 6 summarizes the notation used in the main paper and this supplementary material. Throughout, $\beta < 0$, and $\mathcal{Q}_\beta^{t,s} \subset \mathbb{R}^{t+1,s}$ denotes a pseudo-hyperboloid with t temporal and s spatial tangent dimensions. Accordingly, the ambient space contains $t+1$ temporal coordinates and s spatial coordinates.

Table 6: Notation used in the main paper and supplementary material.

Notation	Description
$\mathbb{R}^{t+1,s}$	Ambient pseudo-Euclidean space with $t+1$ temporal coordinates and s spatial coordinates.
$d = t + s + 1$	Total ambient dimension.
$G = \mathrm{diag}(-I_{t+1}, I_s)$	Metric tensor of $\mathbb{R}^{t+1,s}$.
$\langle x, y\rangle_{\mathcal{Q}} = x^\top G y$	Indefinite inner product in the ambient pseudo-Euclidean space.
$x = (x_t, x_s)$	Temporal–spatial decomposition of an ambient vector, where $x_t \in \mathbb{R}^{t+1}$ and $x_s \in \mathbb{R}^s$.
$\|\xi\|_{\mathcal{Q}} = \sqrt{\lvert\langle \xi, \xi\rangle_{\mathcal{Q}}\rvert}$	Absolute pseudo-norm used for non-null vectors.
$\mathcal{Q}_\beta^{t,s}$	Pseudo-hyperboloid $\{x \in \mathbb{R}^{t+1,s} \mid \langle x, x\rangle_{\mathcal{Q}} = \beta\}$, with $\beta < 0$.
$T_x\mathcal{Q}_\beta^{t,s}$	Tangent space of $\mathcal{Q}_\beta^{t,s}$ at x. The induced metric has signature (t, s).
o	Fixed reference point on $\mathcal{Q}_\beta^{t,s}$ used for pseudo-hyperbolic feature modeling.
$\mathrm{Exp}_x^{\mathcal{Q}}$	Exponential map on $\mathcal{Q}_\beta^{t,s}$.
$\Pi_{\mathcal{Q}}(v)$	Radial projection of a timelike ambient vector v onto $\mathcal{Q}_\beta^{t,s}$.
$h_i^{(0)}$	Initial pseudo-hyperbolic representation of the i-th feature.
s_{ij}	Pseudo-Riemannian similarity score between $h_i^{(0)}$ and $h_j^{(0)}$.
α_{ij}	Attention weight associated with the pair (i, j).
h_i^{att}	Pseudo-hyperbolic representation obtained after attention aggregation and radial projection.
h_i^{res}	Representation obtained after residual interpolation and radial projection.
$a_k,\ b_k$	Euclidean MLR weight and bias for class k.
$w_k = a_k/\|a_k\|_2$	Unit Euclidean class-normal direction.
$\bar{b}_k = b_k/\|a_k\|_2$	Normalized Euclidean class bias.
$H_a^{\mathcal{Q}}$	Pseudo-hyperbolic decision section $\{x \in \mathcal{Q}_\beta^{t,s} \mid \langle x, a\rangle_{\mathcal{Q}} = 0\}$.
$u_k,\ v_k$	Unit spatial and timelike direction used in the generalized Lorentz boost, with $u_k \in \mathbb{R}^s$ and $v_k \in \mathbb{R}^{t+1}$.
ϕ_k	Rapidity parameter $\phi_k = -\bar{b}_k/\sqrt{-\beta}$ used to encode the class-dependent offset.
$B_k(\phi_k)$	Generalized Lorentz boost acting as a pseudo-orthogonal isometry in $\mathbb{R}^{t+1,s}$.
a_k^0	Canonical unit spacelike normal used as the reference normal for class k.
$\tilde{a}_k^s$	Boosted unit spacelike normal, $\tilde{a}_k^s = B_k(\phi_k)a_k^0$.
$H_k^{\mathcal{Q},s}$	Spacelike decision section $\{q \in \mathcal{Q}_\beta^{t,s} \mid \langle q, \tilde{a}_k^s\rangle_{\mathcal{Q}} = 0\}$.
$d_{\mathcal{Q}}^s(p, q)$	Length of an admissible spacelike geodesic connecting p and q.
$d_{\mathcal{Q}}^s(p, H_k^{\mathcal{Q},s})$	Minimum spacelike point-to-hyperplane distance from p to $H_k^{\mathcal{Q},s}$.
$u_k^s(p)$	Spacelike QMLR class score.
$z_k^t(p)$	Surrogate projection used in the timelike branch.
$u_k^t(p)$	Timelike surrogate class score.
$g = \sigma(\gamma)$	Learnable gate controlling the mixture of spacelike and timelike QMLR scores.
$u_k(p)$	Final fused QMLR score for class k.

B Additional Preliminaries

This section provides additional geometric details for the pseudo-hyperboloid used in the main paper. We follow the notation $\mathcal{Q}_\beta^{t,s} \subset \mathbb{R}^{t+1,s}$ with $\beta < 0$, where the ambient space contains $t+1$ temporal coordinates and s spatial coordinates.

B.1 Pseudo-Euclidean Ambient Space

Let $\mathbb{R}^{t+1,s}$ denote $\mathbb{R}^{t+s+1}$ equipped with the indefinite inner product,$G = \mathrm{diag}(-I_{t+1}, I_s)$ is the metric,the inner product of $x, y \in \mathbb{R}^{t+1,s}$ is

$$\langle x, y \rangle_{\mathcal{Q}} = -\sum_{i=0}^{t} x_i y_i + \sum_{j=t+1}^{t+s} x_j y_j = x^\top G y, \tag{17}$$

For a nonzero vector $\xi \in \mathbb{R}^{t+1,s}$, its causal type is determined by $\langle \xi, \xi \rangle_{\mathcal{Q}}$: it is timelike if $\langle \xi, \xi \rangle_{\mathcal{Q}} < 0$, null if $\langle \xi, \xi \rangle_{\mathcal{Q}} = 0$, and spacelike if $\langle \xi, \xi \rangle_{\mathcal{Q}} > 0$. For non-null vectors, we use $\|\xi\|_{\mathcal{Q}} = \sqrt{|\langle \xi, \xi \rangle_{\mathcal{Q}}|}$.

B.2 Pseudo-Hyperboloid and Tangent Geometry

For $\beta < 0$, the pseudo-hyperboloid is defined by

$$\mathcal{Q}_\beta^{t,s} = \left\{ x \in \mathbb{R}^{t+1,s} \mid \langle x, x \rangle_{\mathcal{Q}} = \beta \right\}. \tag{18}$$

Since $\langle x, x \rangle_{\mathcal{Q}} = \beta < 0$, the radial vector x is timelike. Its orthogonal complement therefore induces a pseudo-Riemannian metric of signature (t, s) on $\mathcal{Q}_\beta^{t,s}$.

The tangent space at $x \in \mathcal{Q}_\beta^{t,s}$ is

$$T_x \mathcal{Q}_\beta^{t,s} = \left\{ \xi \in \mathbb{R}^{t+1,s} \mid \langle x, \xi \rangle_{\mathcal{Q}} = 0 \right\}. \tag{19}$$

For an ambient vector $u \in \mathbb{R}^{t+1,s}$, its orthogonal projection onto $T_x \mathcal{Q}_\beta^{t,s}$ is

$$\Pi_x(u) = u - \frac{\langle x, u \rangle_{\mathcal{Q}}}{\beta}\, x. \tag{20}$$

B.3 Geodesics and Exponential Map

Let $\xi \in T_x \mathcal{Q}_\beta^{t,s}$ and define $c = \langle \xi, \xi \rangle_{\mathcal{Q}}$. The geodesic starting from x with initial velocity ξ depends on the causal type of ξ. For $\tau \in \mathbb{R}$,

$$\gamma_{x,\xi}(\tau) = \begin{cases} \cosh\!\left(\dfrac{\tau\sqrt{c}}{\sqrt{-\beta}}\right) x + \sqrt{-\beta}\, \sinh\!\left(\dfrac{\tau\sqrt{c}}{\sqrt{-\beta}}\right) \dfrac{\xi}{\sqrt{c}}, & c > 0, \\ x + \tau\xi, & c = 0, \\ \cos\!\left(\dfrac{\tau\sqrt{-c}}{\sqrt{-\beta}}\right) x + \sqrt{-\beta}\, \sin\!\left(\dfrac{\tau\sqrt{-c}}{\sqrt{-\beta}}\right) \dfrac{\xi}{\sqrt{-c}}, & c < 0. \end{cases} \tag{21}$$

Evaluating the geodesic at $\tau = 1$ gives the exponential map used in the main paper:

$$\mathrm{Exp}_x^{\mathcal{Q}}(\xi) = \begin{cases} \cosh(\rho) x + \sqrt{-\beta}\sinh(\rho) \dfrac{\xi}{\|\xi\|_{\mathcal{Q}}}, & \langle \xi, \xi \rangle_{\mathcal{Q}} > 0, \\ \cos(\rho) x + \sqrt{-\beta}\sin(\rho) \dfrac{\xi}{\|\xi\|_{\mathcal{Q}}}, & \langle \xi, \xi \rangle_{\mathcal{Q}} < 0, \\ x + \xi, & \langle \xi, \xi \rangle_{\mathcal{Q}} = 0, \end{cases} \tag{22}$$

where $\rho = \|\xi\|_{\mathcal{Q}} / \sqrt{-\beta}$.

B.4 Logarithmic Map and Geodesic Connectivity

For $x, y \in \mathcal{Q}_\beta^{t,s}$, define $\alpha = \langle x, y\rangle_{\mathcal{Q}}/\beta$. When x and y are connected by the corresponding non-null geodesic, the logarithmic map can be written as

$$\mathrm{Log}_x^{\mathcal{Q}}(y) = \begin{cases} \dfrac{\operatorname{arcosh}(\alpha)}{\sqrt{\alpha^2-1}}\,(y-\alpha x)\,, & \alpha > 1, \\ \dfrac{\arccos(\alpha)}{\sqrt{1-\alpha^2}}\,(y-\alpha x)\,, & -1 < \alpha < 1. \end{cases} \tag{23}$$

The limiting null case is obtained continuously when the corresponding separation approaches zero.

B.5 Connection to the Lorentz Model

The standard Lorentz model and the pseudo hyperboloid considered in this work arise from the same ambient level-set construction, but their induced manifold metrics are different.

When $t = 0$, the ambient space is $\mathbb{R}^{1,s}$ with the Minkowski metric. Restricting

$$\mathcal{Q}_\beta^{0,s} = \left\{x \in \mathbb{R}^{1,s} \mid \langle x, x\rangle_{\mathcal{Q}} = \beta\right\}$$

yields the standard Lorentz hyperboloid model which satisfies Riemannian condition. Although the ambient metric is indefinite, the metric induced on each tangent space of the hyperboloid is positive definite. As a result, the Lorentz model is a Riemannian manifold.

For $t \geq 1$, the manifold $\mathcal{Q}_\beta^{t,s} \subset \mathbb{R}^{t+1,s}$ has tangent spaces with signature (t, s). Its induced metric is therefore indefinite, and the manifold is pseudo-Riemannian. Accordingly, the geometry studied in this work extends the same hyperboloid-type construction from the Riemannian Lorentz setting to manifolds with multiple temporal directions.

B.6 Admissibility of Pseudo-Hyperbolic Attention

The pairwise condition in Sec. 4.2 is a sufficient condition ensuring that both the attention aggregation and the residual interpolation remain timelike.

Let

$$v_i^{\mathrm{att}} = \sum_j \alpha_{ij} h_j^{(0)}, \qquad \alpha_{ij} \geq 0, \qquad \sum_j \alpha_{ij} = 1.$$

Under the admissibility condition $\langle h_j^{(0)}, h_k^{(0)}\rangle_{\mathcal{Q}} \leq \beta < 0$, we have

$$\begin{aligned} \langle v_i^{\mathrm{att}}, v_i^{\mathrm{att}}\rangle_{\mathcal{Q}} &= \sum_{j,k} \alpha_{ij}\alpha_{ik}\langle h_j^{(0)}, h_k^{(0)}\rangle_{\mathcal{Q}} \\ &\leq \beta \sum_{j,k} \alpha_{ij}\alpha_{ik} = \beta < 0. \end{aligned}$$

Hence, v_i^{att} is timelike and $h_i^{\mathrm{att}} = \Pi_{\mathcal{Q}}(v_i^{\mathrm{att}})$ is well-defined.

Since $\Pi_{\mathcal{Q}}$ rescales by a positive scalar, write $h_i^{\mathrm{att}} = c_i v_i^{\mathrm{att}}$ with $c_i > 0$. Then

$$\langle h_i^{(0)}, h_i^{\mathrm{att}}\rangle_{\mathcal{Q}} = c_i \sum_j \alpha_{ij}\langle h_i^{(0)}, h_j^{(0)}\rangle_{\mathcal{Q}} \leq c_i\beta < 0.$$

For

$$v_i^{\mathrm{res}} = \theta h_i^{(0)} + (1-\theta) h_i^{\mathrm{att}}, \qquad \theta \in [0, 1],$$

we obtain

$$\begin{aligned} \langle v_i^{\mathrm{res}}, v_i^{\mathrm{res}}\rangle_{\mathcal{Q}} = \beta\big(\theta^2 + (1-\theta)^2\big) \\ + 2\theta(1-\theta)\langle h_i^{(0)}, h_i^{\mathrm{att}}\rangle_{\mathcal{Q}} < 0 \end{aligned}$$

for $0 < \theta < 1$, while the endpoints satisfy $\langle v_i^{\mathrm{res}}, v_i^{\mathrm{res}}\rangle_{\mathcal{Q}} = \beta < 0$. Therefore, v_i^{res} is timelike for all $\theta \in [0, 1]$, and the radial projection $h_i^{\mathrm{res}} = \Pi_{\mathcal{Q}}(v_i^{\mathrm{res}})$ is well-defined with

$$\langle h_i^{\mathrm{res}}, h_i^{\mathrm{res}}\rangle_{\mathcal{Q}} = \beta.$$

C Additional Derivations for QMLR

C.1 Euclidean MLR as Signed Distance

For class k, the Euclidean MLR score is $u_k^E(x) = \langle a_k, x\rangle + b_k$, where $a_k \in \mathbb{R}^n$ and $b_k \in \mathbb{R}$. Let $w_k = a_k/\|a_k\|_2$ and $\bar{b}_k = b_k/\|a_k\|_2$. The corresponding decision hyperplane is

$$H_k^E = \left\{x \in \mathbb{R}^n \mid \langle w_k, x\rangle + \bar{b}_k = 0\right\}.$$

Since w_k is a unit Euclidean normal, $d_E(x, H_k^E) = |\langle w_k, x\rangle + \bar{b}_k|$. Hence

$$\begin{aligned} u_k^E(x) &= \|a_k\|_2\big(\langle w_k, x\rangle + \bar{b}_k\big) \\ &= \|a_k\|_2 \ \mathrm{sign}\big(\langle w_k, x\rangle + \bar{b}_k\big) d_E(x, H_k^E). \end{aligned} \tag{24}$$

This signed-distance form motivates QMLR, where the Euclidean hyperplane and distance are replaced by their counterparts on $\mathcal{Q}_\beta^{t,s}$.

C.2 Pseudo-Hyperbolic Decision Sections

For a nonzero ambient normal $a \in \mathbb{R}^{t+1,s}$, we define

$$H_a^{\mathcal{Q}} = \left\{x \in \mathcal{Q}_\beta^{t,s} \mid \langle x, a\rangle_{\mathcal{Q}} = 0\right\}.$$

The causal type of a determines the geometry of the section. For the spacelike branch, we impose $\langle a, a\rangle_{\mathcal{Q}} = 1$. This unit-spacelike constraint is used below to obtain the closed-form spacelike point-to-section distance.

C.3 Generalized Lorentz Boost

A unit spacelike normal determines the orientation of a reference decision section, but does not provide the class-dependent offset represented by the Euclidean bias. We therefore parameterize the normal through a pseudo-orthogonal isometry.

Let $u_k \in \mathbb{R}^s$ and $v_k \in \mathbb{R}^{t+1}$ satisfy $\|u_k\|_2 = \|v_k\|_2 = 1$, and define

$$a_k^0 = \begin{bmatrix} 0_{t+1} \\ u_k \end{bmatrix}, \qquad \langle a_k^0, a_k^0\rangle_{\mathcal{Q}} = 1.$$

With the temporal–spatial ordering used throughout the main paper, the generalized Lorentz boost is

$$B_k(\phi_k) = \begin{bmatrix} I_{t+1} + (\cosh\phi_k - 1)v_k v_k^\top & \sinh\phi_k\, v_k u_k^\top \\ \sinh\phi_k\, u_k v_k^\top & I_s + (\cosh\phi_k - 1)u_k u_k^\top \end{bmatrix}. \tag{25}$$

Proposition 1. $B_k(\phi_k)$ preserves the indefinite metric:

$$B_k(\phi_k)^\top G B_k(\phi_k) = G.$$

Consequently, the causal norm of any transformed vector is preserved.

Proof. The transformation is the identity on the orthogonal complement of the two-dimensional plane spanned by $(v_k, 0)$ and $(0, u_k)$. On this plane it reduces to the standard hyperbolic rotation. Using $\cosh^2\phi_k - \sinh^2\phi_k = 1$ gives $B_k(\phi_k)^\top G B_k(\phi_k) = G$. Therefore $\langle B_k x, B_k y\rangle_{\mathcal{Q}} = \langle x, y\rangle_{\mathcal{Q}}$ for all x, y. □

Following the main paper, the normalized Euclidean offset is encoded by

$$\phi_k = -\frac{\bar{b}_k}{\sqrt{-\beta}},$$

and the class-dependent unit spacelike normal is

$$\tilde{a}_k^s = B_k(\phi_k) a_k^0.$$

By Proposition 1, $\langle \tilde{a}_k^s, \tilde{a}_k^s \rangle_{\mathcal{Q}} = 1$. The corresponding decision section is

$$H_k^{\mathcal{Q},s} = \left\{ q \in \mathcal{Q}_\beta^{t,s} \mid \langle q, \tilde{a}_k^s \rangle_{\mathcal{Q}} = 0 \right\}.$$

Applying $B_k(\phi_k)$ to the canonical normal a_k^0 gives

$$\begin{aligned} \tilde{a}_k^s &= B_k(\phi_k) \begin{bmatrix} 0_{t+1} \\ u_k \end{bmatrix} \\ &= \begin{bmatrix} \sinh(\phi_k)\, v_k \\ \cosh(\phi_k)\, u_k \end{bmatrix}. \end{aligned} \tag{26}$$

Therefore,

$$\begin{aligned} \langle p, \tilde{a}_k^s \rangle_{\mathcal{Q}} &= -\langle p_t, \sinh(\phi_k) v_k \rangle + \langle p_s, \cosh(\phi_k) u_k \rangle \\ &= \cosh(\phi_k) \langle p_s, u_k \rangle - \sinh(\phi_k) \langle p_t, v_k \rangle. \end{aligned} \tag{27}$$

With

$$u_k = \frac{a_{k,s}}{\|a_{k,s}\|_2}, \qquad v_k = \frac{a_{k,t}}{\|a_{k,t}\|_2},$$

the projection term becomes

$$z_k^s(p) = \cosh(\phi_k) \frac{\langle p_s, a_{k,s} \rangle}{\|a_{k,s}\|_2} - \sinh(\phi_k) \frac{\langle p_t, a_{k,t} \rangle}{\|a_{k,t}\|_2}. \tag{28}$$

C.4 Spacelike Point-to-Hyperplane Distance

Let $p \in \mathcal{Q}_\beta^{t,s}$ and write $a = \tilde{a}_k^s$ for brevity. Since $\langle a, a \rangle_{\mathcal{Q}} = 1$, the component of p orthogonal to a is $p_\perp = p - \langle p, a \rangle_{\mathcal{Q}} a$, with

$$\langle p_\perp, p_\perp \rangle_{\mathcal{Q}} = \beta - \langle p, a \rangle_{\mathcal{Q}}^2 < 0.$$

Hence $p_\perp$ is timelike.

We restrict the comparison to an admissible spacelike cluster $\mathcal{A}_k^s(p) \subseteq H_k^{\mathcal{Q},s}$, as defined in Eq. (11). In particular, this cluster contains the normal foot $q^\star$ in Eq. (30), and any two distinct points $q, r \in \mathcal{A}_k^s(p)$ satisfy $\langle q, r \rangle_{\mathcal{Q}} < \beta$. This pairwise requirement is stronger than requiring each candidate endpoint to be spacelike-connected to p. The cluster specifies a mathematical comparison domain, rather than a partition of the input data. The restricted distance is

$$d_{\mathcal{Q}}^s\big(p, H_k^{\mathcal{Q},s}\big) := \inf_{q \in \mathcal{A}_k^s(p)} d_{\mathcal{Q}}^s(p, q). \tag{29}$$

The minimizing endpoint can be obtained explicitly. Define

$$q^\star = \sqrt{\frac{-\beta}{-\beta + \langle p, a \rangle_{\mathcal{Q}}^2}}\, (p - \langle p, a \rangle_{\mathcal{Q}} a)\,. \tag{30}$$

It follows directly that $\langle q^\star, a \rangle_{\mathcal{Q}} = 0$ and $\langle q^\star, q^\star \rangle_{\mathcal{Q}} = \beta$; hence $q^\star \in H_k^{\mathcal{Q},s}$.

Proposition 2. The curve

$$\gamma(\tau) = q^\star \cosh\left(\frac{\tau}{\sqrt{-\beta}}\right) + \sqrt{-\beta}\, a \sinh\left(\frac{\tau}{\sqrt{-\beta}}\right) \tag{31}$$

is a unit-speed spacelike geodesic orthogonal to $H_k^{\mathcal{Q},s}$ at $q^\star$.

Proof. Using $\langle q^\star, q^\star \rangle_{\mathcal{Q}} = \beta$, $\langle q^\star, a \rangle_{\mathcal{Q}} = 0$, and $\langle a, a \rangle_{\mathcal{Q}} = 1$, we obtain $\langle \gamma(\tau), \gamma(\tau) \rangle_{\mathcal{Q}} = \beta$, so the curve remains on $\mathcal{Q}_\beta^{t,s}$. Its derivative satisfies $\langle \dot{\gamma}(\tau), \dot{\gamma}(\tau) \rangle_{\mathcal{Q}} = 1$, hence it is unit-speed and spacelike. At $\tau = 0$, $\dot{\gamma}(0) = a$. For every $\xi \in T_{q^\star} H_k^{\mathcal{Q},s}$, $\langle \xi, a \rangle_{\mathcal{Q}} = 0$, so the geodesic is orthogonal to the decision section. □

Theorem 1 (Minimum spacelike geodesic distance). Let $\mathcal{A}_k^s(p)$ be an admissible spacelike cluster containing $q^\star$. Then $q^\star$ uniquely minimizes $d_Q^s(p,q)$ over $q \in \mathcal{A}_k^s(p)$. Consequently,

$$d_{\mathcal{Q}}^s\big(p, H_k^{\mathcal{Q},s}\big) = d_{\mathcal{Q}}^s(p, q^\star). \tag{32}$$

Proof. For any $q \in H_k^{\mathcal{Q},s}$, $\langle q, a\rangle_{\mathcal{Q}} = 0$, and therefore $\langle p, q\rangle_{\mathcal{Q}} = \langle p_\perp, q\rangle_{\mathcal{Q}}$. Both $p_\perp$ and q are timelike in the orthogonal subspace $a^\perp$. Restricting to the same admissible time cone, the reverse Cauchy–Schwarz inequality gives

$$\langle p_\perp, q\rangle_{\mathcal{Q}} \le -\sqrt{\langle p_\perp, p_\perp\rangle_{\mathcal{Q}}\langle q, q\rangle_{\mathcal{Q}}}.$$

Since $\langle q, q\rangle_{\mathcal{Q}} = \beta$ and $\langle p_\perp, p_\perp\rangle_{\mathcal{Q}} = \beta - \langle p, a\rangle_{\mathcal{Q}}^2$, we obtain

$$\frac{\langle p, q\rangle_{\mathcal{Q}}}{\beta} \ge \sqrt{1 + \frac{\langle p, a\rangle_{\mathcal{Q}}^2}{-\beta}}. \tag{33}$$

Equality holds if and only if q is positively collinear with $p_\perp$ in the same time cone, which gives exactly $q = q^\star$.

For spacelike-geodesically connected points, the geodesic length is

$$d_{\mathcal{Q}}^s(p,q) = \sqrt{-\beta}\,\operatorname{arcosh}\left(\frac{\langle p, q\rangle_{\mathcal{Q}}}{\beta}\right).$$

Since $\operatorname{arcosh}(\cdot)$ is strictly increasing on $[1,\infty)$, the lower bound in equation 33 is attained uniquely at $q^\star$. Thus $q^\star$ minimizes the spacelike geodesic distance over $\mathcal{A}_k^s(p)$. □

Corollary 1. The minimum spacelike point-to-section distance is

$$d_{\mathcal{Q}}^s\big(p, H_k^{\mathcal{Q},s}\big) = \sqrt{-\beta}\left|\operatorname{asinh}\left(\frac{\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}}}{\sqrt{-\beta}}\right)\right|. \tag{34}$$

Proof. From equation 33, equality at $q^\star$ gives

$$\frac{\langle p, q^\star\rangle_{\mathcal{Q}}}{\beta} = \sqrt{1 + \frac{\langle p, a\rangle_{\mathcal{Q}}^2}{-\beta}}.$$

Using $\operatorname{arcosh}(\sqrt{1+x^2}) = |\operatorname{asinh}(x)|$ yields equation 34. Equivalently, if $p = \gamma(r)$ in equation 31, then $\langle p, a\rangle_{\mathcal{Q}} = \sqrt{-\beta}\sinh(r/\sqrt{-\beta})$, so the same result follows from $d_{\mathcal{Q}}^s = |r|$. □

Finally, following the signed-distance form of Euclidean MLR, the spacelike QMLR score is

$$\begin{aligned} u_k^s(p) &= \|a_k\|_2\,\operatorname{sign}\big(\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}}\big) d_{\mathcal{Q}}^s\big(p, H_k^{\mathcal{Q},s}\big) \\ &= \|a_k\|_2\sqrt{-\beta}\,\operatorname{asinh}\left(\frac{\langle p, \tilde{a}_k^s\rangle_{\mathcal{Q}}}{\sqrt{-\beta}}\right), \end{aligned} \tag{35}$$

where we used $\operatorname{sign}(z)|\operatorname{asinh}(z)| = \operatorname{asinh}(z)$.

C.5 Timelike Surrogate Score

The spacelike branch admits a well-defined minimum point-to-section distance because its normal is spacelike. For a timelike normal, the corresponding orthogonal direction is timelike, and the same minimum-distance construction is not available under the indefinite metric. We therefore use the surrogate formulation introduced in the main paper.

Let $z_k^t(p)$ denote the timelike projection term, constructed by exchanging the temporal and spatial roles in the spacelike parameterization:

$$z_k^t(p) = \cosh(\phi_k)\frac{\langle p_t, a_{k,t}\rangle}{\|a_{k,t}\|_2} - \sinh(\phi_k)\frac{\langle p_s, a_{k,s}\rangle}{\|a_{k,s}\|_2}. \tag{36}$$

The timelike class score is then defined as

$$u_k^t(p) = \|a_k\|_2\sqrt{-\beta}\,\operatorname{asinh}\left(\frac{z_k^t(p)}{\sqrt{-\beta}}\right). \tag{37}$$

Unlike u_k^s, u_k^t is not interpreted as a length-minimizing point-to-hyperplane distance. It is a curvature-aware surrogate used to retain a timelike decision branch within the same score scale.

C.6 Adaptive Fusion

The two QMLR branches are combined through a learnable scalar gate. Let $g = \sigma(\gamma) \in (0, 1)$. The final class score is

$$u_k(p) = g\, u_k^t(p) + (1 - g)\, u_k^s(p). \tag{38}$$

The class probabilities are obtained by applying the standard softmax to $\{u_k(p)\}_{k=1}^K$.

C.7 Euclidean Flat Limit

We finally examine the zero-curvature limit $\beta \to -\infty$. Let $c = 1/\sqrt{-\beta}$, so that $c \to 0$. The rapidity satisfies

$$\phi_k = -c\,\bar{b}_k.$$

Using $\cosh(\phi_k) = 1 + O(c^2)$ and $\sinh(\phi_k) = \phi_k + O(c^3)$, the boosted spacelike projection admits a first-order expansion around the reference point of the hyperboloid. Under the standard local Euclidean identification,

$$\langle p, \tilde{a}_k^s \rangle_{\mathcal{Q}} = \langle x, w_k \rangle + \bar{b}_k + O(c). \tag{39}$$

Moreover, $\operatorname{asinh}(cy) = cy + O(c^3)$. Substituting into equation 35 gives

$$\begin{aligned} u_k^s(p) &= \|a_k\|_2 \frac{1}{c} \operatorname{asinh}\left(c\, \langle p, \tilde{a}_k^s \rangle_{\mathcal{Q}}\right) \\ &\longrightarrow \|a_k\|_2 \left(\langle x, w_k \rangle + \bar{b}_k\right) = \langle a_k, x \rangle + b_k. \end{aligned} \tag{40}$$

Thus the spacelike QMLR branch recovers the Euclidean MLR score in the flat limit.

C.8 Relation to Existing Hyperbolic MLRs

Existing hyperbolic MLR formulations are typically defined either through tangent-space hyperplanes or through horospherical constructions. QMLR instead parameterizes class-dependent decision sections directly in the ambient pseudo-Euclidean space. The spacelike branch admits the closed-form minimum distance derived above, while the timelike branch is retained through a surrogate score and combined adaptively through equation 38.

D Experimental Details and Additional Results

D.1 EEG Cross-Session Decoding

D.1.1 Datasets and Evaluation Protocol

We evaluate Q-Net on three cross-session EEG benchmarks covering motor imagery (MI), steady-state visual evoked potentials (SSVEP), and error-related negativity (ERN). The datasets are BCI Competition IV 2a (BCIC-IV-2a), MAMEM-SSVEP-II, and the BCI Challenge ERN dataset, respectively.

Motor imagery. BCIC-IV-2a contains EEG recordings from nine subjects performing four motor-imagery tasks: left hand, right hand, feet, and tongue. Each subject participated in two recording sessions conducted on different days. Each session contains 288 trials, with 72 trials for each class. The first session is used for model development, and the second session is used for cross-session evaluation. A subset of the training session is held out for validation and model selection.

SSVEP. MAMEM-SSVEP-II contains recordings from eleven subjects under five visual stimulation frequencies: 6.66, 7.50, 8.57, 10.00, and 12.00 Hz. Each subject contains five recording sessions. Following the cross-session setting, the earlier sessions are used for model development and the final session is reserved for cross-session evaluation.

ERN. The BCI-ERN dataset contains EEG recordings collected during a P300-based spelling task. The problem is formulated as binary classification between correct and error-related responses. Because the two classes are imbalanced, we report the area under the ROC curve (AUC) for evaluation.

Accuracy is reported for MI, while AUC is reported for SSVEP and ERN. Each experiment is repeated ten times, and the results are reported as mean $\pm$ standard deviation. For each run, the checkpoint with the lowest validation loss is used for evaluation.

Table 7: Dimensional transformations of Q-Net on the MI, SSVEP, and ERN benchmarks. The manifold representations are processed by the pseudo-hyperbolic attention module described in the main paper.

Stage	**MI**	**SSVEP**	**ERN**
Input	$1 \times 22 \times 438$	$1 \times 8 \times 125$	$1 \times 56 \times 160$
Temporal Conv.	$64 \times 22 \times 438$	$64 \times 8 \times 125$	$64 \times 56 \times 160$
Temporal Kernel	$(1, 15)$	$(1, 50)$	$(1, 25)$
Temporal Padding	$(0, 7)$	$(0, 25)$	$(0, 12)$
Spatial Conv.	$64 \times 1 \times 438$	$64 \times 1 \times 125$	$64 \times 1 \times 160$
Spatial Kernel	$(22, 1)$	$(8, 1)$	$(56, 1)$
Average Pooling	$64 \times 1 \times 25$	$64 \times 1 \times 4$	$64 \times 1 \times 6$
Euclidean Sequence	64×25	64×4	64×6
Pseudo-Hyperbolic Embedding	65×25	65×4	65×6
Pseudo-Hyperbolic Attention	65×25	65×4	65×6
Aggregated Representation	65	65	65
Output Classes	4	5	2

D.1.2 Network Architecture

Q-Net follows a Euclidean-to-pseudo-hyperbolic pipeline. A shallow convolutional encoder first extracts temporal and spatial EEG features. The resulting sequence is interpreted as tangent-space features and mapped to $\mathcal{Q}_\beta^{t,s}$ through the exponential map. Pseudo-hyperbolic attention is then applied to refine interactions among the manifold representations, followed by aggregation and the proposed QMLR classifier.

The dimensional transformations for the three datasets are summarized in Table 7.

D.1.3 Training Details

All models are optimized using cross-entropy loss. The dataset-specific training hyperparameters are listed in Table 8. The curvature scale β and timelike dimension are treated as dataset-specific geometric hyperparameters. All experiments use the same evaluation protocol as described above.

D.1.4 Additional Ablation Studies

We further examine the design choices in the pseudo-hyperbolic attention and residual refinement modules on BCIC-IV-2a. These experiments isolate the effects of the geometric similarity used by attention and the manifold-preserving residual update.

Attention formulation. We compare the proposed pseudo-Riemannian attention with three alternatives. *Ambient Euclidean* replaces the pseudo-Riemannian inner product with a standard Euclidean

Table 8: Hyperparameter settings for EEG cross-session decoding.

Hyperparameter	**MI**	**SSVEP**	**ERN**
Epochs	300	100	150
CNN Channels	64	64	64
Batch Size	128	64	128
Learning Rate	5×10^{-3}	5×10^{-3}	1×10^{-2}
Weight Decay	0.45	0.45	0.45
timelike Dimension	20	40	20
β	-3.0	-2.0	-1.0
Dropout	0.4	0.3	0.4

inner product in the ambient space. *Tangent Euclidean* computes attention after mapping the representations to the tangent space. *Raw pseudo-Riemannian* retains the proposed pseudo-Riemannian attention scores but removes the subsequent attention aggregation, allowing us to distinguish the contribution of geometric similarity from that of feature aggregation.

As shown in Table 10, the pseudo-Riemannian formulation gives the strongest result among the three simplified variants. Both ambient and tangent-space Euclidean attention perform slightly worse, suggesting that the indefinite pseudo-Riemannian similarity is better aligned with the geometry of the learned representations. The comparison with the raw pseudo-Riemannian variant further separates the geometric affinity from the subsequent aggregation operation.

Residual refinement. We next ablate the residual update following pseudo-hyperbolic attention. *Off* removes the residual connection entirely. *No projection* performs the ambient residual interpolation without projecting the result back to the pseudo-hyperboloid. *Tangent* performs the residual operation in the tangent space. *Fixed* retains the proposed ambient residual and projection but replaces the learnable interpolation coefficient with a fixed coefficient.

Table 11 shows that tangent-space residual refinement performs substantially worse than the ambient-space alternatives. Removing the manifold projection also provides no clear advantage over directly disabling the residual connection. Among the simplified variants, the fixed ambient residual performs best, indicating that residual refinement is most effective when carried out in the ambient pseudo-Euclidean space while respecting the manifold constraint through projection. The learnable interpolation used in the full model further allows this balance to be adapted during training.

D.1.5 Additional Visualizations

We provide additional qualitative analyses of the learned EEG representations. These visualizations are used only for post-hoc inspection and are not involved in model training or model selection.

Figure 5 compares the temporal and spatial responses of Q-Net for correct and error trials. The error condition shows more sustained responses over fronto-central channels, whereas the correct condition exhibits a more concentrated temporal pattern. The scalp maps also show stronger contributions around central and fronto-central electrodes. These observations provide a qualitative view of the class-dependent spatiotemporal structure captured by the model.

D.2 Image Classification

D.2.1 Datasets and Setup

We evaluate QMLR on CIFAR-10 and CIFAR-100, which contain $60{,}000$ RGB images of size 32×32 with 10 and 100 classes, respectively. We use the standard split of $50{,}000$ training images and $10{,}000$ test images.

Following the protocol used in the main paper, all methods share the same ResNet-18 backbone and differ only in the final classification head. The Euclidean features produced by the backbone are mapped to the corresponding manifold before applying the geometry-specific classifier. We compare

Table 9: Hyperparameters for CIFAR-10 and CIFAR-100.

Hyperparameter	**CIFAR-10**	**CIFAR-100**
Epochs	300	300
Batch size	128	128
Learning rate	0.1	0.1
Weight decay	5×10^{-4}	5×10^{-4}
Optimizer	SGD	SGD
Precision	32-bit	32-bit
β	-1	-1
Timelike dimension t	2	2

Table 10: Ablation of the attention formulation on BCIC-IV-2a. Results are reported as accuracy (%).

Attention Variant	Balanced Acc.
Ambient Euclidean	71.98 ± 1.25
Tangent Euclidean	71.84 ± 1.30
Raw pseudo-Riemannian	$\mathbf{72.18 \pm 1.21}$

Table 11: Ablation of the residual refinement on BCIC-IV-2a. Results are reported as balanced accuracy (%).

Residual Variant	Balanced Acc.
Off	72.95 ± 1.03
No projection	73.03 ± 1.16
Tangent	70.13 ± 1.63
Fixed	$\mathbf{73.16 \pm 1.02}$

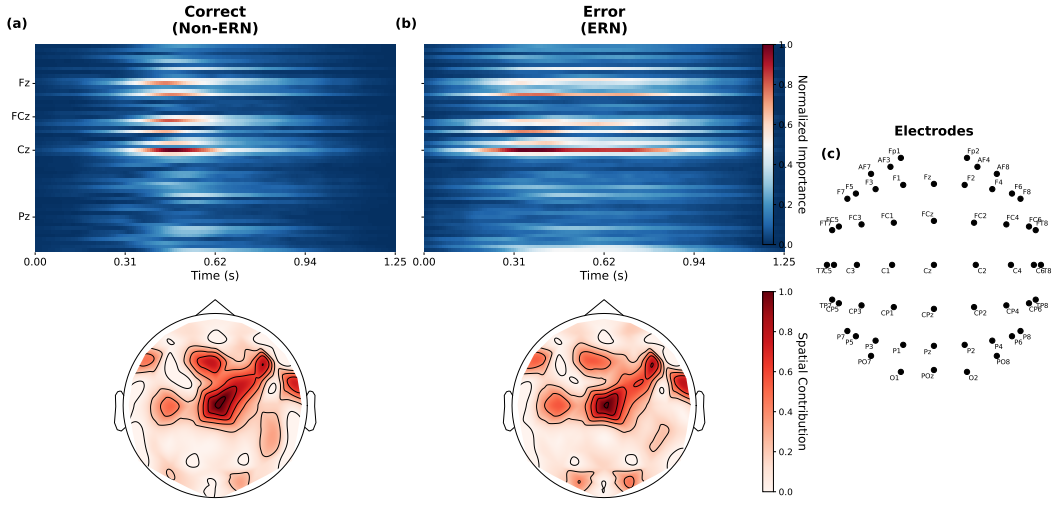


Figure 5: Additional visualization of Q-Net on the BCI-ERN dataset. (a) Correct temporal saliency (b) Error temporal saliency (c) Electrode configuration All values are normalized independently for visualization.

Euclidean MLR, Poincaré MLR (PMLR), Pseudo-Busemann MLR (PBMLR), Busemann Poincaré MLR (BMLR-P), Lorentz MLR (LMLR), Busemann Lorentz MLR (BMLR-L), and the proposed QMLR. The baseline results follow the corresponding experimental protocol used in the main paper.

D.2.2 Training Details

All models are trained with cross-entropy loss. For QMLR, the curvature parameter is fixed to $\beta = -1$ and the pseudo-hyperbolic timelike dimension is set to $t = 2$. The remaining optimization settings are summarized in Table 9.

D.3 Genome Sequence Learning

D.3.1 Datasets

We evaluate QMLR on the Transposable Elements Benchmark (TEB) and selected tasks from the Genome Understanding Evaluation (GUE) benchmark.

TEB. TEB contains seven binary sequence-classification datasets covering plant retrotransposons, plant DNA transposons, and human pseudogenes. The adopted datasets are LTR Copia, LINEs, SINEs, CMC-EnSpm, hAT-Ac, processed pseudogenes, and unprocessed pseudogenes. We follow the chromosome-level training, validation, and test splits provided by the benchmark. Dataset statistics are summarized in Table 12.

GUE. GUE contains genomic sequence tasks from multiple species and biological settings. Following the experiments in the main paper, we use the core promoter detection, promoter detection, COVID-variant classification, and transcription-factor prediction subsets. The adopted dataset statistics are provided in Table 13.

Table 12: Summary statistics for TEB.

Task	Species	Dataset	Classes	Max Length	Train / Dev / Test
Retrotransposons	Plant	LTR Copia	2	500	7666 / 682 / 568
		LINEs		1000	22502 / 2030 / 1782
		SINEs		500	21152 / 1836 / 1784
DNA transposons	Plant	CMC-EnSpm	2	200	19912 / 1872 / 1808
		hAT-Ac		1000	17322 / 1822 / 1428
Pseudogenes	Human	processed	2	1000	17956 / 1046 / 1740
		unprocessed		1000	12938 / 766 / 884

D.3.2 Additional Experiment

We further evaluate QMLR on the Mouse Transcription Factor Prediction tasks. All experimental settings follow those described in the main genomic experiment.

E Limitations and Future Work

Although the proposed framework demonstrates the feasibility of learning with pseudo-Riemannian representations across several tasks, the potential of pseudo-Riemannian geometry remains far from fully explored. The present work focuses primarily on representation modeling, pseudo-hyperbolic attention, and QMLR, while many geometric operations and network components are still implemented through relatively simple constructions. Developing a broader family of intrinsic pseudo-Riemannian operators and studying their behavior in deeper architectures remain important directions.

Another limitation lies in the timelike branch of QMLR. Unlike the spacelike branch, which admits a closed-form point-to-hyperplane distance under the admissible spacelike geometry, the current timelike formulation is introduced as a surrogate score. It provides a practical way to retain timelike decision behavior, but does not yet admit the same minimum-distance interpretation. A more

Table 13: Summary statistics for the adopted GUE datasets.

Task	Species	Dataset	Cls.	L.	Train / Dev / Test
Core promoter detection	Human	tata	2	70	4904 / 613 / 613
		notata			42452 / 5307 / 5307
		all			47356 / 5920 / 5920
Promoter detection	Human	tata	2	300	4904 / 613 / 613
		notata			42452 / 5307 / 5307
		all			47356 / 5920 / 5920
COVID variant classification	Virus	Covid	9	1000	77669 / 7000 / 7000
Transcription Factor prediction	Mouse	Ch12Nrf2Iggrab(0)	2	100	6478 / 810 / 810
		Ch12Znf384hpa004051 Iggrab(1)			53952 / 6745 / 6745
		MelJundIggrab(2)			2620 / 328 / 328
		MelMafkDm2p5dStd(3)			1904 / 239 / 239
		MelNelfeIggrab(4)			15064 / 1883 / 1883

Table 14: Additional genomic results on Mouse Transcription Factor Prediction. Performance is measured by Matthews correlation coefficient (MCC).

Dataset	MLR	PMLR	LMLR	BMLR-L	QMLR
0	22.51±2.78	**46.10±1.92**	42.90±2.35	44.77±2.13	43.75±1.75
1	76.56±0.51	79.22±0.53	**79.93±0.69**	78.90±0.64	78.23±0.82
2	62.69±1.52	67.03±2.56	64.04±3.55	65.06±2.75	**70.91±2.91**
3	36.93±8.35	63.16±4.99	56.06±1.95	57.11±2.03	**65.06±2.94**
4	30.23±3.13	41.40±0.83	**42.28±1.26**	41.41±2.04	41.41±1.12

principled timelike decision formulation with a rigorous geometric characterization is therefore an important direction for future work.

Future studies may also investigate adaptive learning of the pseudo-Riemannian signature and causal structure instead of treating the temporal and spatial dimensions as predefined hyperparameters. In addition, extending the current framework to larger-scale models, deeper manifold networks, and more diverse multimodal or structured data may further clarify when and how pseudo-Riemannian geometry provides useful inductive biases.