Hyperbolic Representation Learning for Longitudinal Medical Imaging

Taha Emre
Institute of Artificial Intelligence
Medical University of Vienna
taha.emre@meduniwien.ac.at

Thomas Pinetz
Institute of Artificial Intelligence
Medical University of Vienna

Hrvoje Bogunović
Institute of Artificial Intelligence
Medical University of Vienna

Abstract

Longitudinal 3D medical imaging captures disease trajectories over time and is central to monitoring progression in many diseases. Existing self-supervised learning (SSL) methods for volumetric medical images largely operate in Euclidean latent spaces and do not explicitly encode temporal ordering. We propose a self-supervised framework that learns longitudinal representations in Lorentz hyperbolic space. For visit pairs from the same patient at different times, our model combines a hyperbolic contrastive objective with a temporal entailment loss and a time difference aware hyperbolic margin loss to enforce progression-consistent ordering in the time-like (radial) dimension. Experiments on MRI- and OCT-based degenerative disease cohorts show strong performance versus recent Euclidean and hyperbolic baselines, and ablations confirm that temporal constraints are important. To the best of our knowledge, this is the first hyperbolic representation learning approach for longitudinal medical imaging trajectories.

1 Introduction

Longitudinal 3D scans (e.g., serial MRI and OCT) provide time-resolved structural information that is essential for tracking disease onset and progression. Supervised approaches that exploit longitudinal labels achieve strong performance but require expensive annotations and curated follow-up data. Self-supervised learning (SSL) offers a way to pretrain robust volumetric encoders from unlabelled imaging trajectories and then fine-tune for downstream clinical tasks, reducing annotation demands.

Most prior SSL methods for volumetric imaging extend contrastive or reconstruction objectives and operate on Euclidean embeddings. Euclidean latent spaces, however, lack intrinsic mechanisms to represent directed progression: distances grow linearly and do not naturally capture hierarchical divergence or accelerating change. In particular, they do not possess an inherent directionality that can be used for temporal alignment. On the other hand, hyperbolic geometry, characterized by constant negative curvature and exponential volume growth with radius, has been successful for hierarchical and tree-like data and has recently been adapted for representation learning in vision. In particular, the Lorentz (hyperboloid) model provides a natural time-like coordinate: distance from the origin increases exponentially and can encode progression along a preferred direction. We hypothesize that hyperbolic latent spaces provide useful inductive biases for longitudinal trajectories where small changes near baseline correspond to slowly evolving states while larger radii capture accelerated divergence. Based on this hypothesis, we propose to couple contrastive learning with hyperbolic temporal ordering for longitudinal representations.

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

2 Related work

Longitudinal SSL Siamese-based approaches for longitudinal data typically use pairs or short subsequences with pretext tasks such as time-gap regression [10], order prediction [8], or temporal-change prediction within contrastive learning [4]. Another line of work focuses on time-related directionality using a learnable temporal embedding [11] or metadata-aware kernels [3].

Hyperbolic representation learning Hyperbolic representation learning aims to capture inherent data structures, such as whole-part relations [6, 1], specific-vs-prototype behavior, and hierarchical organization. Most methods extract embeddings in Euclidean tangent space and then lift them to hyperbolic space for loss computation and evaluation. MERU [2], a seminal method, learns image-text representations in hyperbolic space by modeling cross-modal hierarchy: text embeddings are placed closer to the origin (more generic) and image embeddings farther away (more specific), with an entailment loss that enforces image points to lie inside text-defined cones. In medical imaging, hyperbolic methods have mainly targeted semantic hierarchies [1] or anomaly detection [5], rather than explicit longitudinal temporal ordering.

3 Method

Problem A patient in a longitudinal medical imaging cohort is regularly monitored over multiple visits. A 3D scan of a patient at visit time t is denoted by $\mathbf{x}_t \in \mathbb{R}^{D\times H\times W}$. A transformer-based encoder $f_\theta : \mathcal{X} \to \mathbb{L}^n_K$ maps each scan to a hyperbolic representation $\mathbf{z}_t = f_\theta(\mathbf{x}_t)$ on the Lorentz hyperboloid with curvature $-K$. The downstream predictor h_ψ estimates conversion within horizon τ from a single-visit embedding $\mathbf{z}_t$.

Hyperbolic Spaces Hyperbolic space is a complete, simply connected Riemannian manifold $(\mathcal{M}, g)$ with constant sectional curvature $-K$ $(K > 0)$, yielding exponential volume growth relative to Euclidean space. Among equivalent models (Poincaré ball, upper half-space, and Lorentz/hyperboloid), we use the Lorentz model because it is numerically stable at large radii and admits closed-form geodesics [7]. In the Lorentz model, $\mathbb{L}^n_K$ is embedded in Minkowski space $\mathbb{R}^{n+1}$ with Lorentzian inner product:

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

The manifold constraint is $\langle \mathbf{x}, \mathbf{x}\rangle_{\mathcal{L}} = -1/K$. The n-dimensional hyperboloid of curvature $-K$ is the upper sheet:

$$\mathbb{L}^n_K = \left\{ \mathbf{x} \in \mathbb{R}^{n+1} \mid \langle \mathbf{x}, \mathbf{x}\rangle_{\mathcal{L}} = -\frac{1}{K},\ x_0 > 0 \right\}, \tag{2}$$

The tangent space at a point $\mathbf{z} \in \mathbb{L}^n_K$ is a Euclidean space consisting of vectors Lorentz-orthogonal to that point:

$$T_{\mathbf{z}}\mathbb{L}^n_K = \{\mathbf{v} \in \mathbb{R}^{n+1} \mid \langle \mathbf{z}, \mathbf{v}\rangle_{\mathcal{L}} = 0\}. \tag{3}$$

The exponential map lifts $\mathbf{v} \in T_{\mathbf{x}}\mathbb{L}^n_K$ to the manifold:

$$\exp_{\mathbf{x}}(\mathbf{v}) = \cosh\left(\sqrt{K}\,\|\mathbf{v}\|_{\mathcal{L}}\right)\mathbf{x} + \sinh\left(\sqrt{K}\,\|\mathbf{v}\|_{\mathcal{L}}\right)\frac{\mathbf{v}}{\sqrt{K}\,\|\mathbf{v}\|_{\mathcal{L}}}. \tag{4}$$

The origin is defined as $\mathbf{o} = \left(1/\sqrt{K}, 0, \ldots, 0\right)$. For $\mathbf{x} = (x_{time}, x_{space})$ with $x_{space} \in \mathbb{R}^n$, the time-like coordinate is recovered from the spatial coordinates as:

$$x_{time} = \sqrt{\frac{1}{K} + \|x_{space}\|^2}, \tag{5}$$

which reconstructs the full Lorentz point from its spatial part [2, 9].

In MERU [2], the encoder backbone produces image embeddings in Euclidean tangent space around the hyperboloid origin, allowing standard Vision Transformer architectures. These embeddings are then lifted with the exponential map to obtain hyperbolic embeddings $\mathbf{z}$ for pretraining and inference.

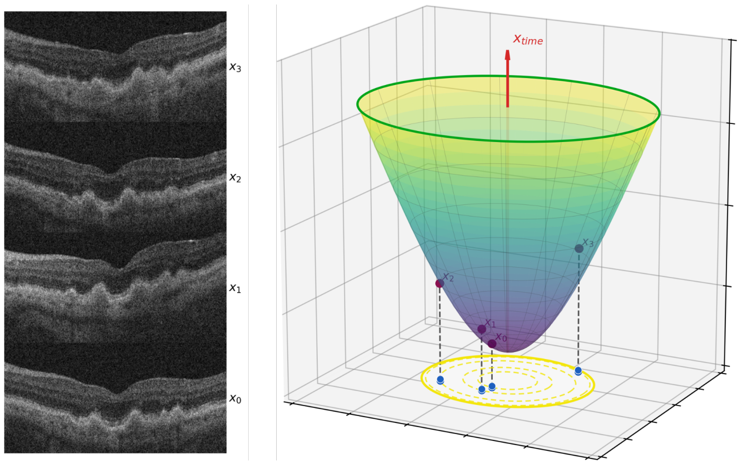

Figure 1: Our proposed models aims to utilize inherit time like dimension x_{time} of Lorentz model to learn correct temporal ordering of medical images.

3.1 Temporal entailment loss

Earlier-visit scans are inherently more ambiguous (prototype-like) and should lie closer to the hyperbolic origin; later visits should be placed at larger radii along a consistent direction. The underlying assumption is that as degenerative diseases progress, images become more specific. We use the time-like dimension z_{time} to enforce longitudinal ordering in hyperbolic space as $z_{t+\Delta t,time} > z_{t,time}$. In our formulation, this is aligned with radial ordering because, in the Lorentz model, z_{time} is a monotonic function of radius (equivalently, distance to the origin).

We adapt the cone-based MERU loss to a longitudinal setup. For a sampled pair of visits at times t and $t + \Delta t$ of one subject, with $\Delta t > 0$, let $\mathbf{z}_t = f_\theta(\mathbf{x}_t)$ and $\mathbf{z}_{t+\Delta t} = f_\theta(\mathbf{x}_{t+\Delta t})$. A cone half-aperture at $\mathbf{z}_t$ that depends on its spatial norm (preventing degenerate cones near the origin) is defined as:

$$\mathrm{aper}(\mathbf{z}_t) = \arcsin\left(\frac{2\varepsilon}{\sqrt{K}\,\|\mathbf{z}_{t,space}\|}\right), \qquad \varepsilon > 0. \tag{6}$$

The exterior angle $\mathrm{ext}(\mathbf{z}_t, \mathbf{z}_{t+\Delta t})$ measures angular deviation of $\mathbf{z}_{t+\Delta t}$ relative to the cone axis at $\mathbf{z}_t$. It is defined as:

$$\mathrm{ext}(\mathbf{z}_t, \mathbf{z}_{t+\Delta t}) = \cos^{-1}\left(\frac{z_{t+\Delta t,time} + z_{t,time}\cdot(K\langle \mathbf{z}_t, \mathbf{z}_{t+\Delta t}\rangle_{\mathcal{L}})}{\|\mathbf{z}_{t,space}\|\sqrt{(K\langle \mathbf{z}_t, \mathbf{z}_{t+\Delta t}\rangle_{\mathcal{L}})^2 - 1}}\right) \tag{7}$$

The temporal entailment loss is $\mathcal{L}_{\text{entail}}(\mathbf{z}_{t+\Delta t}; \mathbf{z}_t) = \max\big(0,\ \mathrm{ext}(\mathbf{z}_t, \mathbf{z}_{t+\Delta t}) - \mathrm{aper}(\mathbf{z}_t)\big)$. The final training objective combines the contrastive term with the entailment term:

$$\mathcal{L}_{meru} = \mathcal{L}_{\text{cont}} + \lambda\,\mathcal{L}_{\text{entail}} \tag{8}$$

For pull/push operations, we use the geodesic distance in the Lorentz model:

$$d_{\mathbb{L}}(x, y) = \frac{1}{\sqrt{K}} \cdot \operatorname{arccosh}\left(-K\langle x, y\rangle_{\mathcal{L}}\right) \tag{9}$$

3.2 Hyperbolic Margin Loss

While MERU loss (Eq. 8) enforces temporal ordering, it does not explicitly account for varying Δt. Additionally, the contrastive term pulls $\mathbf{z}_t$ and $\mathbf{z}_{t+\Delta t}$ closer. We therefore introduce a margin-based loss based on Δt and distance to the origin, focusing on progression in the time-like direction:

$$\mathcal{L}_{\text{margin}} = \max\left(0, \operatorname{arccosh}\left(1 + \frac{\Delta t}{K}\right) + d_{\mathbb{L}}(\mathbf{o}, \mathbf{z}_t) - d_{\mathbb{L}}(\mathbf{o}, \mathbf{z}_{t+\Delta t})\right) \tag{10}$$

The final objective is $\mathcal{L}_{\text{margin}} + \mathcal{L}_{meru}$.

Table 1: ADNI results.

Method	1 year		3 years	
	PRAUC	ROCAUC	PRAUC	ROCAUC
yAware [3]	0.185	0.725	0.285	**0.717**
ATMG [9]	0.123	0.657	0.232	0.654
MERU [2]	0.193	**0.733**	0.289	0.709
Ours	**0.240**	0.708	**0.327**	0.704

Table 2: HARBOR 12-month results.

Method	PRAUC	ROCAUC
TC [4]	**0.123**	**0.605**
ATMG [9]	0.104	0.576
MERU [2]	0.117	0.602
Ours	0.113	0.586

4 Experiments

We used longitudinal cohorts including ADNI (brain MRI) and HARBOR (retinal OCT), each with multiple visits per patient, for pretraining and downstream evaluation. During pretraining, we fixed curvature to $K = 1$ and scaled the final layer normalization to avoid numerical overflow. Linear evaluation uses clinical conversion tasks within specific time windows (ADNI: 1y and 3y conversion from Mild Cognitive Impairment to Alzheimer's disease; HARBOR: 12-month conversion from intermediate age-related macular degeneration (AMD) to wet-AMD) from a single visit. We report PRAUC and ROCAUC. The PRAUC baselines are 0.07 for ADNI 1y, 0.18 for ADNI 3y, and 0.11 for HARBOR 12m. For hyperbolic models, classification is performed in hyperbolic space using Lorentz multinomial logistic regression. Comparisons include hyperbolic baselines MERU [2] and ATMG [9], and domain-specific longitudinal baselines yAware [3] (ADNI) and TC [4] (HARBOR).

Table 3: ADNI ablation on temporal constraints (PRAUC).

Method	ADNI 1y	ADNI 3y
Ours	**0.240**	**0.327**
Hyperbolic Contrastive	0.132	0.304
Reversed Temporal Order	0.142	0.245

Results Our model improves minority-class performance (PRAUC), while ROCAUC changes are more modest (Table 1). This is expected in imbalanced longitudinal tasks, where detecting future converters is harder than global ranking. HARBOR remains a more challenging regime in terms of ROCAUC, where performance differences are minimal across models (Table 2).

Ablation Table 3 isolates temporal design choices. Removing entailment and margin loss terms hurts PRAUC, indicating the need for temporal SSL. Additionally, we reverse temporal direction to test whether healthier scans are less specific. The performance drop supports the assumption that our approach benefits from the inherent hierarchical structure of hyperbolic space and from the inductive bias that progression should map to outward radial movement.

5 Conclusion

Positioning longitudinal embeddings in Lorentz hyperbolic space and enforcing temporal entailment together with a margin constraint yields intrinsic ordering of visits without auxiliary information. To the best of our knowledge, this is the first application of hyperbolic representation learning to longitudinal medical imaging trajectories. We also observe that, due to the exponential geometry of hyperbolic space, the margin term can induce large embedding norms, which may cause numerical overflow and degraded performance. This suggests that the margin definition requires further tuning. As the next step, future work will directly leverage the intrinsic time-like dimension for risk estimation tasks.

References

[1] Baek, S., Lee, J., Sim, J., Jeong, M., Kim, W.H.: Mnm: Multi-level neuroimaging meta-analysis with hyperbolic brain-text representations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 397–406. Springer (2025)

[2] Desai, K., Nickel, M., Rajpurohit, T., Johnson, J., Vedantam, S.R.: Hyperbolic image-text representations. In: International Conference on Machine Learning. pp. 7694–7731. PMLR (2023)

[3] Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., Favre, P., Polosan, M., Mcdonald, C., Piguet, C.M., et al.: Contrastive learning with continuous proxy meta-data for 3d mri classification. In: Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. pp. 58–68. Springer (2021)

[4] Emre, T., Chakravarty, A., Lachinov, D., Rivail, A., Schmidt-Erfurth, U., Bogunović, H.: Learning temporally equivariance for degenerative disease progression in OCT by predicting future representations. In: Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. pp. 196–206. Springer (2024)

[5] Gonzalez-Jimenez, A., Lionetti, S., Amruthalingam, L., Gottfrois, P., Gröger, F., Pouly, M., Navarini, A.A.: Is hyperbolic space all you need for medical anomaly detection? In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 312–322. Springer (2025)

[6] Hsu, J., Gu, J., Wu, G., Chiu, W., Yeung, S.: Capturing implicit hierarchical structure in 3d biomedical images with self-supervised hyperbolic representations. Advances in neural information processing systems **34**, 5112–5123 (2021)

[7] Ibrahimi, S., Atigh, M.G., Noord, N.V., Mettes, P., Worring, M.: Intriguing properties of hyperbolic embeddings in vision-language models. Transactions on Machine Learning Research (2024), https://openreview.net/forum?id=P5D2gfi4Gg

[8] Kim, H., Sabuncu, M.R.: Learning to compare longitudinal images. In: Medical Imaging with Deep Learning. pp. 350–367. PMLR (2023)

[9] Ramasinghe, S., Shevchenko, V., Avraham, G., Thalaiyasingam, A.: Accept the modality gap: An exploration in the hyperbolic space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27263–27272 (2024)

[10] Rivail, A., Schmidt-Erfurth, U., Vogl, W.D., Waldstein, S.M., Riedl, S., Grechenig, C., Wu, Z., Bogunovic, H.: Modeling disease progression in retinal oCTs with longitudinal self-supervised learning. In: International Workshop on PRedictive Intelligence In MEdicine. pp. 44–52. Springer (2019)

[11] Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. Medical Image Analysis **71**, 102051 (2021). https://doi.org/10.1016/j.media.2021.102051