Angular-Constrained Hyperbolic Learning for Hierarchical Multimodal Survival Prediction

Haotian Yang*, Qing Zhang*, Qingli Li, and Yan Wang(✉)

Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China
ywang@cee.ecnu.edu.cn

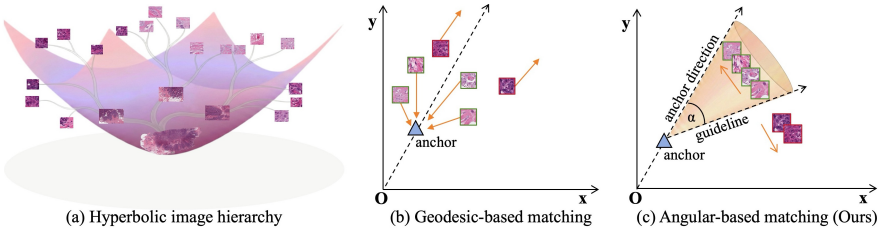
Abstract. Multimodal learning has demonstrated strong potential for cancer survival prediction through the joint modeling of whole-slide images (WSIs) and genomic data. However, existing approaches struggle to capture deep interactions between genomic signals and multiscale pathological representations. Although hyperbolic geometry naturally accommodates hierarchical structures, most current hyperbolic multimodal methods still adopt proximity-driven contrastive paradigms and rely primarily on geodesic distance for similarity estimation. Such designs overemphasize cross-modal spatial proximity while neglecting intrinsic modality discrepancies, thereby weakening modality-specific structures and compromising hierarchical organization. To overcome these limitations, we propose HyperGP, an angular-constrained hyperbolic learning framework for multimodal survival prediction. It explicitly models pathology as a patch–region–slide hierarchy and derives prognostic anchors from pathway-level genomics. We introduce a geometry-aware hierarchical representation module to embed both modalities into a shared hyperbolic manifold and obtain risk-specific prototypes. These embeddings and prototypes are subsequently aligned by a designed unified multimodal angular contrastive learning module while preserving modality-specific structures. Finally, we propose an angular hyperbolic-guided survival prediction module for precise survival forecasting. Extensive experiments across multiple cancer cohorts demonstrate the superior performance of the proposed HyperGP, highlighting the effectiveness of angular constraints for hierarchical survival modeling in hyperbolic space. Code is available at: https://github.com/yang1987-star/hyperGP.

Keywords: Computational pathology · Genomics · Hyperbolic learning · Multimodal survival prediction

1 Introduction

Cancer survival prediction plays a pivotal role in clinical decision-making, as it guides treatment strategies [22, 2]. However, accurate survival prediction remains

H. Yang and Q. Zhang—Denotes equal contribution.



(a) Hyperbolic image hierarchy (b) Geodesic-based matching (c) Angular-based matching (Ours)

Fig. 1. Hyperbolic pathology hierarchy (a), and a comparison of geodesic-distance (b) versus our angular-constrained matching objectives in hyperbolic space (c).

a challenging task due to the complex nature of cancer and its heterogeneous manifestation across patients. Therefore, multimodal learning has been emerging for survival prediction by integrating multiple sources, such as whole slide images (WSIs), genomic data, and clinical variables [1, 9, 24, 27]. Genomic data, with its ability to capture molecular features, complements the structure information provided by WSIs. It offers more comprehensive understanding of the underlying biology [2, 9, 25]. Despite utilizing these diverse data, existing multimodal methods still struggle to fully capture the complex interactions between genomic signals and multiscale pathological features.

Most multimodal approaches follow a prevailing "tokenize-and-fuse" paradigm, where features from WSIs and genomic data are tokenized and then simply fused via attention mechanisms [1, 24, 9, 27]. While these methods are successful in the survival prediction task, they often overlook the hierarchical semantic relationships between different data modalities. Specifically, traditional multimodal methods typically rely on distance metrics, such as Euclidean distance, to align multi-level features within a single modality or across modalities [24, 27, 26, 3]. Although distance metrics are effective at capturing geometric proximity, they fail to model the deep semantic differences within hierarchical structures. As a result, they struggle to capture the complex relationships that exist between features at different levels, especially in the context of survival prediction.

In contrast, hyperbolic geometry offers a natural solution for embedding the hierarchical structures, where the depth of hierarchical structures is often represented by the distance to the root [7, 23]. This approach is particularly useful for capturing the inherent hierarchies present in both genomic and pathological data. However, most existing methods overemphasize cross-modal spatial proximity and rely heavily on geodesic distances, which neglect the intrinsic discrepancies between modalities [13, 17]. This results in weakened modality-specific structures and compromised representation of the hierarchical relationships that are crucial for accurate prediction. Therefore, while hyperbolic geometry offers a promising solution, it is essential to incorporate more effective methods that preserve both the spatial relationships and structural integrity of each modality.

To address these challenges, we propose an angular-constrained hyperbolic learning framework to preserve modality-specific structure and hierarchy-aware cross-modal relations, named as HyperGP. Specifically, a geometry-aware hier-

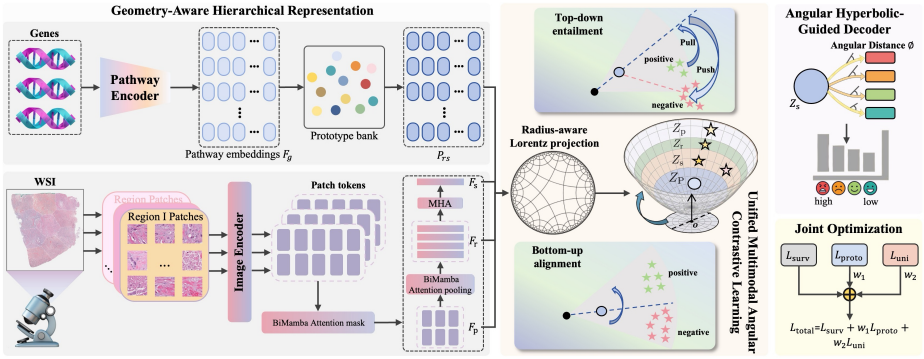


Fig. 2. Overview of HyperGP: WSIs are hierarchically compressed from patches to regions to slides, genomics is distilled into prognostic anchors. Both modalities are embedded into a hyperbolic space by geometry-aware hierarchical representation block, and aligned by unified multimodal angular contrastive learning block. The aligned features are input into an angular hyperbolic-guided decoder for survival prediction.

archical representation module is introduced to embed hierarchical pathological features (patch-region-slide) and pathway-level genomic representations into one shared hyperbolic space, enabling the construction of risk-specific prototypes. Subsequently, we design a novel unified multimodal angular contrastive learning module to align these embeddings and prototypes. This angular formulation ensures hierarchy-aware alignment between the modalities while maintaining the intrinsic structures unique to each modality. Finally, we propose an angular hyperbolic-guided survival prediction module utilizing the aligned embeddings and the prototypes. Extensive experiments across five TCGA cohorts demonstrate that angular-guided multimodal learning strategy in hyperbolic space outperforms the state-of-the-art (SOTA) unimodal and multimodal methods.

2 Methodology

2.1 Preliminaries

Lorentz Model. We embed representations in the Lorentz model $\mathbb{L}_\kappa^d$, a Riemannian manifold with constant negative curvature $-\kappa$ ($\kappa > 0$). It is defined as $\mathbb{L}_\kappa^d = \{z \in \mathbb{R}^{d+1} \mid \langle z, z\rangle_\mathcal{L} = -1/\kappa,\ z_0 > 0\}$, where $z = (z_0, z_{1:d})$, $z_0 = \sqrt{1/\kappa + \|z_{1:d}\|_2^2}$, and $\langle u, v\rangle_\mathcal{L} = -u_0 v_0 + u_{1:d}^\top v_{1:d}$ denotes the Lorentzian inner product [14]. Tangent vectors η at the origin $o = (1/\sqrt{\kappa}, 0, \ldots, 0)$ are projected to $\mathbb{L}_\kappa^d$ via the exponential map $\exp_o(\eta) = \cosh(\|\eta\|_\mathcal{L})o + \eta \sinh(\|\eta\|_\mathcal{L})/\|\eta\|_\mathcal{L}$, where $\|\cdot\|_\mathcal{L}$ is the Lorentzian norm.

Directed Angular Entailment. To model hierarchies, we adopt the exterior angle $\mathrm{ext}(u, v)$ from entailment cones using the Euclidean norm of the spatial component $\|u_{1:d}\|_2$ [4, 23]. Based on this, we define the top-down entailment score $\mathcal{E}^{\mathrm{td}}(u, v) = \pi - \mathrm{ext}(u, v)$ and the symmetric angular distance $\phi(u, v) =$

$\text{ext}(u,v)+\text{ext}(v,u)-\pi$. Maximizing $\mathcal{E}^{\text{td}}(u,v)$ encourages v to lie inside the cone of u, while ϕ serves as a robust symmetric similarity metric for prediction.

$$\text{ext}(u,v)=\arccos\left(\frac{v_0+u_0\kappa\langle u,v\rangle_{\mathcal{L}}}{\|u_{1:d}\|_2\sqrt{(\kappa\langle u,v\rangle_{\mathcal{L}})^2-1}}\right). \tag{1}$$

2.2 Geometry-Aware Hierarchical Representation

As illustrated in Fig. 2, HyperGP establishes a unified hierarchical structure spanning genomic anchors, slides, regions, and patches. Nodes are instantiated at each semantic level and embedded into the Lorentz manifold $\mathbb{L}_\kappa^d$ with radius-aware scaling. Each WSI is tiled into small patches and these patches are encoded by a frozen CTransPath encoder [21]. The patch features F_{p} are spatially partitioned into G groups, and the top-K patches in each group are maintained by calculating importance scores through an attention block. These selected patch features F_{p} are aggregated into region-level features F_{r} via attention pooling, and further pooled into a slide-level feature F_{s}. This bottom-up aggregation produces a multi-scale pathological representation set $\{F_{\text{p}}, F_{\text{r}}, F_{\text{s}}\}$.

For genomic data, we follow PIBD [25] to obtain the patient-level genomic embedding F_{g} by a pathway-wise self-normalizing network, and then get the corresponding risk-specific prototype $P_{\text{rs}}=\{p_{a,c}\mid a\in\{0,1\},c\in\{1,\cdots,C\}\}$ by prototypical information bottleneck. Here, a represents if the patient is dead and C represents all time bins. The genomic embedding F_{g} is encouraged to match its risk-specific prototype P_{rs}, and the learned prototypes serve as the genomic anchors in the hyperbolic space. For the i^{th} patient, we obtain the genomic embedding F_{g_i}, the matched risk-specific prototype p_i. F_{g_i} is aligned to the risk-specific prototype P_{rs} via a Gaussian-regularized matching loss $\mathcal{L}_{\text{proto}}$ [25]:

$$\mathcal{L}_{\text{proto}}=\frac{1}{B}\sum_{i=1}^{B}\left[-\text{Sim}(F_{g_i},p_i)+\frac{1}{2C-1}\sum_{p_{a,c}\neq p_i}\text{Sim}(F_{g_i},p_{a,c})\right], \tag{2}$$

where B is the batch size, $\text{Sim}(\cdot,\cdot)$ denotes the cosine similarity function. During optimization, the genomic embedding is closer to its corresponding true prototype while repelled from others. To mirror the exponential volume expansion of hierarchies, we project all the above node types $\nu\in\{F_{\text{p}},F_{\text{r}},F_{\text{s}},P_{\text{rs}}\}$ into the Lorentz manifold using type-specific learnable scalars α_ν:

$$z_\nu=\exp_o(\alpha_\nu\cdot h_\nu), \tag{3}$$

where $h_\nu\in\mathbb{R}^d$ is the corresponding Euclidean feature derived from the hierarchical features. By learning the scaling factors α_ν, the network automatically positions abstract concepts (risk-specific prototypes) close to the origin and concrete instances (patches) nearer to the boundary, thereby faithfully preserving the intrinsic geometric depth across modalities.

2.3 Unified Multimodal Angular Contrastive Learning

While hyperbolic projection provides the representations with a geometry well suited to hierarchical data, it does not guarantee the desired parent-child semantics across scales. Therefore, we introduce a bidirectional angular contrastive learning over three adjacent parent-child relations $\mathcal{R} = \{rs\_s, s\_r, r\_p\}$, corresponding to risk-specific prototype-to-slide, slide-to-region, and region-to-patch, respectively. For each relation $r \in \mathcal{R}$, let (u, v) denote the parent and child embeddings in hyperbolic space. We explicitly enforce both top-down angular entailment and bottom-up angular alignment. The angular entailment ensures that the child feature lies within the parent's outward cone, which is measured by the angular distance in hyperbolic space: $\mathcal{E}_r^{\mathrm{td}}(u, v) = \pi - \mathrm{ext}(u, v)$. Conversely, the angular alignment constrains the parent to align with the child's inward (i.e., toward-origin) direction, preventing parent nodes from drifting away from their descendants. This constraint is measured as: $\mathcal{E}_r^{\mathrm{bu}}(v, u) = \mathrm{ext}(v, u)$.

Let $\mathcal{P}_r^{\mathrm{td}}(u)$ denote the set of positive child nodes associated with a parent node u under relation r. To avoid cross-level semantic ambiguity, we restrict the candidate set $\mathcal{V}_r^{\mathrm{td}}(u)$ to all child nodes at the same hierarchical level within the current mini-batch. The top-down InfoNCE loss [5] is formulated as:

$$\mathcal{L}_r^{\mathrm{td}}(u) = -\log \frac{\sum_{v \in \mathcal{P}_r^{\mathrm{td}}(u)} \exp(\mathcal{E}^{\mathrm{td}}(u, v)/\tau)}{\sum_{v' \in \mathcal{V}_r^{\mathrm{td}}(u)} \exp(\mathcal{E}^{\mathrm{td}}(u, v')/\tau)}. \tag{4}$$

where τ is a shared temperature for both directions. Conversely, let $\mathcal{P}_r^{\mathrm{bu}}(v)$ denote the set of positive parent nodes associated with a child node v under relation r, and $\mathcal{V}_r^{\mathrm{bu}}(v)$ represent the candidate set of all same-level parent nodes within one mini-batch. The bottom-up loss is formulated as:

$$\mathcal{L}_r^{\mathrm{bu}}(v) = -\log \frac{\sum_{u \in \mathcal{P}_r^{\mathrm{bu}}(v)} \exp(\mathcal{E}^{\mathrm{bu}}(v, u)/\tau)}{\sum_{u' \in \mathcal{V}_r^{\mathrm{bu}}(v)} \exp(\mathcal{E}^{\mathrm{bu}}(v, u')/\tau)}. \tag{5}$$

where τ is a shared temperature parameter used for both directions. The total loss for this unified asymmetric angular contrastive learning module is:

$$\mathcal{L}_{\mathrm{uni}} = \sum_{r \in \mathcal{R}} w_r \big(\mathbb{E}_u[\mathcal{L}_r^{\mathrm{td}}(u)] + \gamma_r \, \mathbb{E}_v[\mathcal{L}_r^{\mathrm{bu}}(v)] \big), \tag{6}$$

where $\mathbb{E}$ represents mathematical expectation operation, w_r and γ_r are hyperparameters that control the weights. This decoupled formulation maintains the depth-wise modality structure while preventing cross-level semantic ambiguity.

2.4 Angular Hyperbolic-Guided Decoder and Optimization

To maintain the integrity of the learned geometry, we perform survival prediction directly on the hyperbolic manifold. The follow-up time is partitioned into C bins. For the i^{th} patient, $z_{\mathrm{s}_i} \in \mathbb{L}_\kappa^d$ denote the hyperbolic embedding of

Table 1. Performance comparison using C-index (mean ± std) on five TCGA cohorts. Best results are highlighted in bold and the second-best in underlined.

	Method	BLCA	COADREAD	HNSC	STAD	BRCA	Overall
Gene	MLP [6]	0.530±0.077	0.712±0.114	0.520±0.064	0.497±0.031	0.622±0.079	0.576
	SNN [10]	0.521±0.070	0.711±0.162	0.514±0.076	0.485±0.047	0.621±0.073	0.570
	SNNTrans [10]	0.583±0.060	0.739±0.124	0.570±0.035	0.547±0.041	0.679±0.053	0.622
Pathology	ABMIL [8]	0.624±0.061	0.727±0.118	0.602±0.028	0.629±0.029	0.632±0.066	0.643
	CLAMMB [12]	0.618±0.077	0.743±0.142	0.616±0.043	0.617±0.036	0.645±0.086	0.648
	DSMIL [11]	0.644±0.069	0.693±0.142	0.580±0.019	0.640±0.046	0.660±0.047	0.643
	TransMIL [15]	0.630±0.012	0.627±0.074	0.615±0.039	0.586±0.018	0.629±0.026	0.617
	RRTMIL [19]	0.630±0.042	0.696±0.125	0.592±0.020	0.651±0.031	0.651±0.059	0.644
	MHIMMIL [18]	0.614±0.070	0.717±0.142	0.609±0.023	0.641±0.052	0.643±0.052	0.645
Multimodal	Porpoise [2]	0.617±0.056	0.738±0.151	0.614±0.058	<u>0.660</u>±0.106	0.668±0.070	0.660
	MCAT [1]	0.640±0.076	0.724±0.137	0.564±0.084	0.625±0.118	0.685±0.109	0.647
	MOTCat [24]	0.659±0.069	0.742±0.124	0.582±0.041	0.621±0.065	0.727±0.027	0.681
	SurvPath [9]	0.649±0.074	0.751±0.149	0.612±0.063	0.643±0.039	0.712±0.084	0.673
	PIBD [25]	<u>0.661</u>±0.041	<u>0.786</u>±0.111	0.619±0.056	0.652±0.063	0.726±0.062	<u>0.689</u>
	MMP [16]	0.661±0.037	0.691±0.081	0.598±0.045	0.649±0.069	<u>0.731</u>±0.053	0.666
	CCL [28]	0.659±0.062	0.773±0.131	<u>0.621</u>±0.074	0.640±0.067	0.691±0.066	0.677
	HyperGP	**0.671**±0.021	**0.845**±0.092	**0.637**±0.062	**0.672**±0.059	**0.732**±0.069	**0.711**

their slide-level pathological feature F_{s_i}, obtained via Eq. 3. Similarly, we get $z_{p_{1,c}} \in \mathbb{L}_\kappa^d$ of $p_{1,c}$, i.e., the abstract concept of a death event ($a = 1$) occurring at time bin c. We compute the hazard logit using the symmetric angular distance $\phi(\cdot, \cdot)$ between z_{s_i} and $z_{p_{1,c}}$. Then, the distance is converted into the discrete hazard by a sigmoid operation, denoted as $\lambda_{i,c}$. The angular distance emphasizes semantic direction and primarily encodes hierarchical depth, making it particularly suitable for matching a concrete instance (slide) to an abstract concept (risk-specific prototype). We optimize the survival prediction block using the discrete-time negative log-likelihood:

$$\mathcal{L}_{\text{surv}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j=1}^{\bar{c}-1}\log(1-\lambda_{i,j}) + \bar{a}\log\lambda_{i,\bar{c}} + (1-\bar{a})\log(1-\lambda_{i,\bar{c}})\right], \quad (7)$$

where N is the total number of samples, $\bar{a}$ is the true death status, and $\bar{c}$ is the true time-bin index of the i^{th} patient. The total loss function for HyperGP is a weighted sum of the above loss functions $\{\mathcal{L}_{\text{surv}}, \mathcal{L}_{\text{proto}}, \mathcal{L}_{\text{uni}}\}$:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{surv}} + w_1\mathcal{L}_{\text{proto}} + w_2\mathcal{L}_{\text{uni}}, \quad (8)$$

where w_1 and w_2 are hyperparameters balancing the contributions of auxiliary tasks. This joint optimization ensures that the learned features are discriminative for survival prediction and consistent with the intrinsic hierarchical structure of biological concepts, enhancing generalization and interpretability.

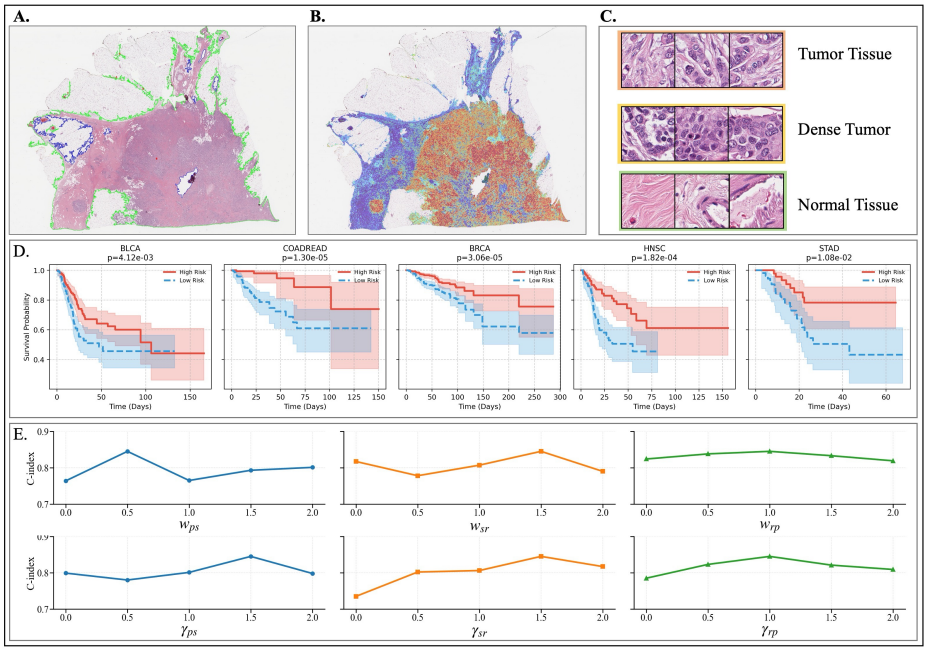


Fig. 3. Visualization of tissue-type localization map (A), model-derived risk heatmap on WSI regions (B), and representative patch crops of pathological categories (C). (D) Kaplan–Meier curves across five TCGA cohorts. (E) Evaluation of coefficients in HyperGP on COADREAD.

3 Experiments

3.1 Experimental Settings

Datasets and Implementation Details. We evaluate on five TCGA cohorts [20]: BRCA ($n = 968$), BLCA ($n = 384$), COADREAD ($n = 298$), STAD ($n = 317$), and HNSC ($n = 392$). Each patient includes a diagnostic WSI, matched RNA-seq profile, survival time, and censoring indicator. WSIs are tiled into 224×224 patches at 20× magnification and encoded via CTransPath [21]. To reduce background noise, patches are partitioned into $G = 10$ groups, retaining $K = 80\%$ per group. Genomic pathway tokens are derived from Reactome and MSigDB. Continuous survival times are partitioned into $C = 4$ quartile-based discrete bins [25]. We employ 5-fold cross-validation (4:1 split) and report the C-index. Models are implemented in PyTorch on a single RTX 3090 GPU and optimized via Adam (batch size 32, learning rate 1×10^{-4}, weight decay 1×10^{-5}). The hierarchical loss hyperparameters for prototype-to-slide, slide-to-region, and region-to-patch relations are $(w_{ps}, w_{sr}, w_{rp}) = (0.5, 1.5, 1.0)$ and $(\gamma_{ps}, \gamma_{sr}, \gamma_{rp}) = (1.5, 1.5, 1.0)$, with $\tau = 0.08$.

Table 2. Ablation study on component effectiveness.

type-specific α	UMACL	AHGD	BLCA	COADREAD
✗	✗	✗	0.624±0.023	0.713±0.084
✓	✗	✗	0.631±0.045	0.737±0.091
✓	✓	✗	0.659±0.057	0.794±0.104
✓	✓	✓	**0.671**±0.021	**0.845**±0.092

3.2 Comparison with SOTA Methods

We compare HyperGP with SOTA methods under pathology-only, genomics-only, and multimodal settings. As shown in Table 1, HyperGP achieves the best overall C-index of 0.711, outperforming all unimodal and multimodal methods on nearly all datasets. Notably, it ranks first across all five cohorts, demonstrating consistent superiority. The most pronounced improvement is observed on COADREAD, where HyperGP attains 0.845 compared to 0.786 for the best competing method. While multimodal methods generally excel by leveraging complementary signals, competitive unimodal performance on COADREAD (CLAMMB: 0.743) underscores the challenge of modality discrepancy. HyperGP's consistent superiority proves that hierarchy-aware angular alignment effectively stabilizes cross-modal learning. As shown in Fig. 3 (A)-(C), attention heatmaps successfully localize high-risk pathological patterns. Furthermore, Kaplan-Meier analysis (Fig. 3 (D)) confirms significant survival stratification between high- and low-risk groups across all cohorts ($p < 0.05$).

3.3 Ablation Study

Ablation Analysis. Table 2 details the ablation results on BLCA and COADREAD. With patch filtering and PIBD-based prototype construction fixed, we evaluate three key components. First, type-specific α better preserves hierarchy-aware radial organization in hyperbolic space. Second, the Unified Multimodal Angular Contrastive Learning (UMACL) enforces directional parent-child constraints across prototype–slide–region–patch levels, promoting consistent multi-level alignment. Finally, the Angular Hyperbolic-Guided Decoder (AHGD) replaces the Euclidean decoder, ensuring the final risk prediction is geometry-consistent with the learned hyperbolic hierarchy. Jointly optimizing these components yields the most discriminative multimodal representations. We further replace angular relations with geodesic matching, which decreases the overall C-index from 0.711 to 0.673. And replacing the Lorentz hyperbolic space with a Euclidean angular control gives 0.689, indicating that both hyperbolic geometry and angular relation modeling contribute to the final performance.

Hyperparameter Analysis. Fig. 3 (E) analyzes the UMACL coefficients on COADREAD. The selected directional factors $(\gamma_{\text{ps}}, \gamma_{\text{sr}}, \gamma_{\text{rp}})$ and level weights $(w_{\text{ps}}, w_{\text{sr}}, w_{\text{rp}})$ optimally balance the top-down and bottom-up hierarchical supervision. This configuration minimizes cross-level semantic ambiguity and ensures robust hyperbolic alignment.

4 Conclusion

In this paper, we proposed HyperGP, a framework that treats multimodal survival prediction as a hierarchical geometric alignment problem. By shifting from Euclidean feature fusion to angular-constrained hyperbolic learning, HyperGP unifies genomic signals and multi-scale pathology in a shared Lorentz manifold. The designed unified multimodal angular contrastive learning module integrates with the angular hyperbolic-guided decoder, enforcing a directed hierarchy from genomic anchors to slides, regions, and patches, ensuring that prognostic prediction remains consistent with hierarchical biological inclusion. Experiments on five TCGA cohorts show that hierarchy-aware angular supervision in hyperbolic space provides an effective inductive bias for robust multimodal prognostication.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Grant Nos. 62471182, 62475072, and 42527804), the Shanghai Rising-Star Program (Grant No. 24QA2702100), the Fundamental Research Funds for the Central Universities, the Science and Technology Commission of Shanghai Municipality Basic Research Program (Grant No. 25JD1401300), the Science and Technology Commission of Shanghai Municipality (Grant Nos. 25xtcx00600 and 22DZ2229004), the Joint Laboratory of Hyperspectral Big Data and Artificial Intelligence (No. MIP202602), and the AI Project from the Economic and Information Commission of Shanghai (Grant No. 2024-GZL-RGZN-01038).

Disclosure of Interests. The authors have no competing interests to declare in relation to this paper.

References

1. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4025 (2021)
2. Chen, R.J., Lu, M.Y., Williamson, D.F., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., et al.: Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer cell **40**(8), 865–878 (2022)
3. Ding, T., Wagner, S.J., Song, A.H., Chen, R.J., Lu, M.Y., Zhang, A., Vaidya, A.J., Jaume, G., Shaban, M., Kim, A., et al.: A multimodal whole-slide foundation model for pathology. Nature medicine pp. 1–13 (2025)
4. Ganea, O., Bécigneul, G., Hofmann, T.: Hyperbolic entailment cones for learning hierarchical embeddings. In: International conference on machine learning. pp. 1646–1655. PMLR (2018)
5. Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 297–304. JMLR Workshop and Conference Proceedings (2010)
6. Haykin, S.: Neural networks: a comprehensive foundation. Prentice hall PTR (1994)

7. Huang, P., Huang, Y., Zhao, W., He, J., Yu, L.: Hyperpath: Knowledge-guided hyperbolic semantic hierarchy modeling for wsi analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 262–272. Springer (2025)
8. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
9. Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11579–11590 (2024)
10. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. Advances in neural information processing systems **30** (2017)
11. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
12. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
13. Mandica, P., Franco, L., Kallidromitis, K., Petryk, S., Galasso, F.: Hyperbolic learning with multimodal large language models. In: European Conference on Computer Vision. pp. 382–398. Springer (2024)
14. Nickel, M., Kiela, D.: Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In: International conference on machine learning. pp. 3779–3788. PMLR (2018)
15. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
16. Song, A.H., Chen, R.J., Jaume, G., Vaidya, A.J., Baras, A.S., Mahmood, F.: Multimodal prototyping for cancer survival prediction. arXiv preprint arXiv:2407.00224 (2024)
17. Srivastava, S., Wu, K.: Hypervlm: Hyperbolic space guided vision language modeling for hierarchical multi-modal understanding. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2347–2358 (2025)
18. Tang, W., Huang, S., Zhang, X., Zhou, F., Zhang, Y., Liu, B.: Multiple instance learning framework with masked hard instance mining for whole slide image classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4078–4087 (2023)
19. Tang, W., Zhou, F., Huang, S., Zhu, X., Zhang, Y., Liu, B.: Feature re-embedding: Towards foundation model-level performance in computational pathology. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11343–11352 (2024)
20. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia **2015**(1), 68–77 (2015)
21. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
22. Wang, Z., Zhang, Y., Xu, Y., Imoto, S., Chen, H., Song, J.: Histo-genomic knowledge association for cancer prognosis from histopathology whole slide images. IEEE Transactions on Medical Imaging **44**(5), 2170–2181 (2025)

23. Wang, Z., Ramasinghe, S., Xu, C., Monteil, J., Bazzani, L., Ajanthan, T.: Learning visual hierarchies in hyperbolic space for image retrieval. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9924–9934 (2025)
24. Xu, Y., Chen, H.: Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21241–21251 (2023)
25. Zhang, Y., Xu, Y., Chen, J., Xie, F., Chen, H.: Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction. arXiv preprint arXiv:2401.01646 (2024)
26. Zhao, H., Tao, D., Zhan, Y., Ni, J., Chen, Y.: Cpgnet: Multimodal graph learning with hierarchical category guidance for multi-label whole slide image classification. IEEE Journal of Biomedical and Health Informatics (2025)
27. Zhou, F., Chen, H.: Cross-modal translation and alignment for survival analysis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21485–21494 (2023)
28. Zhou, H., Zhou, F., Chen, H.: Cohort-individual cooperative learning for multimodal cancer survival analysis. IEEE Transactions on Medical Imaging **44**(2), 656–667 (2024)