HyperPath: Knowledge-Guided Hyperbolic Semantic Hierarchy Modeling for WSI Analysis

Peixiang Huang[1], Yanyan Huang[1], Weiqin Zhao[1], Junjun He[2,3], and Lequan Yu[1] (✉)

[1] The University of Hong Kong, Hong Kong SAR, China
{paxson_huang, yanyanh, wqzhao98}@connect.hku.hk, lqyu@hku.hk
[2] Shanghai Artificial Intelligence Laboratory, Shanghai, China
[3] Shanghai Innovation Institute, Shanghai, China
hejunjun@sjtu.edu.cn

Abstract. Pathology is essential for cancer diagnosis, with multiple instance learning (MIL) widely used for whole slide image (WSI) analysis. WSIs exhibit a natural hierarchy—patches, regions, and slides—with distinct semantic associations. While some methods attempt to leverage this hierarchy for improved representation, they predominantly rely on Euclidean embeddings, which struggle to fully capture semantic hierarchies. To address this limitation, we propose **HyperPath**, a novel method that integrates knowledge from textual descriptions to guide the modeling of semantic hierarchies of WSIs in hyperbolic space, thereby enhancing WSI classification. Our approach adapts both visual and textual features extracted by pathology vision-language foundation models to the hyperbolic space. We design an Angular Modality Alignment Loss to ensure robust cross-modal alignment, while a Semantic Hierarchy Consistency Loss further refines feature hierarchies through entailment and contradiction relationships and thus enhance semantic coherence. The classification is performed with geodesic distance, which measures the similarity between entities in the hyperbolic semantic hierarchy. This eliminates the need for linear classifiers and enables a geometry-aware approach to WSI analysis. Extensive experiments show that our method achieves superior performance across tasks compared to existing methods, highlighting the potential of hyperbolic embeddings for WSI analysis. The source code is available at https://github.com/HKU-MedAI/HyperPath.

Keywords: Hierarchical Representation Learning · Hyperbolic Space · Vision-Language Model · Whole Slide Image.

1 Introduction

Pathology is the gold standard for cancer diagnosis, and whole slide image (WSI) analysis is a key component of computational pathology, advancing cancer diagnosis and prognosis through machine learning [14]. However, due to the large size and complex patterns of WSIs, pixel-level annotations are impractical. Multiple Instance Learning (MIL) [20] addresses this by operating on bags of image

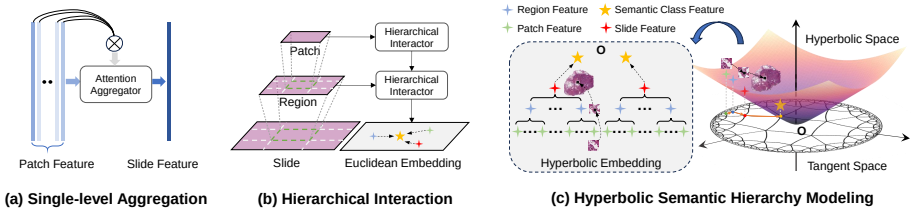

Fig. 1. Comparison of different representation learning approaches for WSI.

patches without exhaustive labeling, enabling slide-level representation learning for downstream tasks. Some attention-based MIL methods [11,19,33] leverage aggregation operators to combine patch-level information, providing interpretable and effective representations. TransMIL [24] incorporates Transformers to aggregate morphological and spatial features efficiently, while DTFD-MIL [32] introduces pseudo-bags to address challenges posed by small sample sizes.

Despite these advantages, simple single-level feature aggregation often fails to explicitly model the hierarchical structure of slides. Hierarchical modeling is essential as it captures both local details and global context by representing complex relationships across hierarchical levels. To address this, methods like [2,8,10] extract multi-scale features to model spatial hierarchies. However, these may not fully preserve intrinsic semantic hierarchies. This limitation has motivated exploration into hyperbolic modeling, a paradigm well-suited for hierarchical structures. Recent studies [5,15,22,23,26,29,16] have demonstrated its effectiveness, especially in capturing visual and textual hierarchical relationships. In WSIs, hyperbolic modeling can organize levels (patch-region-slide) to align with semantic and hierarchical structures, as more intuitively shown in Fig. 1.

In this paper, we introduce **HyperPath**, a novel method that leverages textual concept knowledge to model hierarchical semantic relationships in WSIs within hyperbolic space, improving classification performance. Building on the success of foundation models in pathology [3,9,18,27,28], we employ CONCH [18] for feature extraction. Specifically, our framework encodes image patches and class description prompts into hyperbolic space, where hierarchical aggregation is performed to extract region- and slide-level features that reflect the intrinsic structure of WSIs. By leveraging geodesic distances between slide representations and semantic class features, our method achieves robust WSI classification without relying on linear classifiers. To enhance the semantic hierarchy in hyperbolic space, we propose two key loss functions: **Angular Modality Alignment Loss** and **Semantic Hierarchy Consistency Loss**. The former minimizes cross-modality discrepancies, ensuring effective and robust alignment across hierarchical levels, while the latter enforces semantic coherence within and across modalities, addressing contradictions and promoting entailments. Extensive experiments on four TCGA tasks demonstrate the effectiveness of our approach and its ability to learn semantic hierarchies in WSIs.

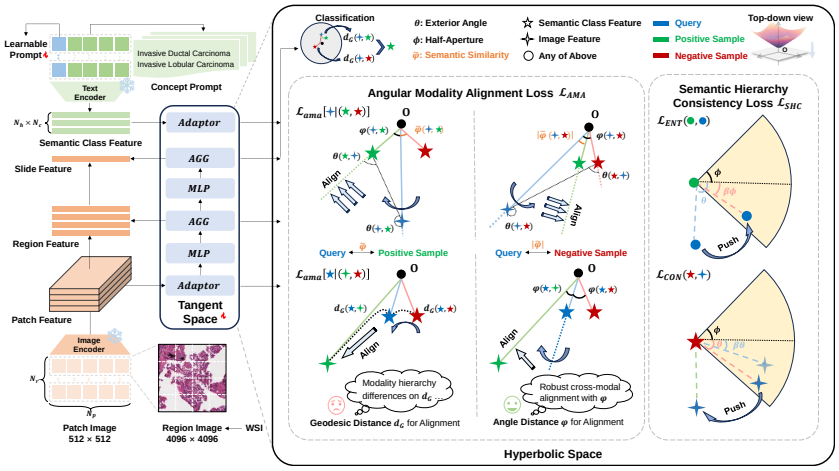

Fig. 2. Overview of our proposed HyperPath framework. The WSI images are hierarchically aggregated, simultaneously optimized in hyperbolic space. Guided by semantic class feature extracted from textual concepts, we utilize Angular Modality Alignment Loss and Semantic Hierarchy Consistency Loss to learn semantic hierarchies in WSIs.

2 Methodology

2.1 Preliminaries

Hyperbolic Space. Hyperbolic geometry exhibits exponential space expansion, allowing it to naturally represent hierarchical structures with efficient scaling, so that it can embed complex relationships without excessive distortion [12,25,31]. Following [5], we choose the Lorentz model to present the k-dimensional hyperbolic space with curvature $-\rho$, denoted by $\mathbb{H}^k_\rho$, due to its numerical stability and efficiency. It can be described by $(k+1)$-dimensional Euclidean space $\mathbb{R}^{k+1}$, where for every vector $\mathbf{u} \in \mathbb{R}^{k+1}$, the first dimension corresponds to the time component $\mathbf{u}_t \in \mathbb{R}$, the remaining dimensions represent the space component $\mathbf{u}_s \in \mathbb{R}^k$, satisfying $\mathbf{u}_t = \sqrt{1/\rho + \|\mathbf{u}_s\|^2_{\mathbb{E}}}$, where $\|\cdot\|_{\mathbb{E}}$ is the Euclidean norm. Let the Euclidean and Lorentzian inner product be denoted as $\langle\cdot,\cdot\rangle_{\mathbb{E}}$ and $\langle\cdot,\cdot\rangle_{\mathbb{H}}$ respectively, they satisfy the following equation $\langle \mathbf{u}, \mathbf{v}\rangle_{\mathbb{H}} = \langle \mathbf{u}_s, \mathbf{v}_s\rangle_{\mathbb{E}} - \mathbf{u}_t\mathbf{v}_t$. Thus the hyperbolic space can be defined as $\mathbb{H}^k_\rho = \{\mathbf{u} \in \mathbb{R}^{k+1} : \langle \mathbf{u}, \mathbf{u}\rangle_{\mathbb{H}} = -1/\rho,\ \rho > 0\}$ and the induced Lorentzian norm can be expressed as $\|\mathbf{u}\|_{\mathbb{H}} = \sqrt{|\langle \mathbf{u}, \mathbf{u}\rangle_{\mathbb{H}}|}$.

Tangent Space. The tangent space is an orthogonal Euclidean space linked to each point in hyperbolic space, allowing projections that preserve hyperbolic geometry. The origin $\mathbf{O}$ is commonly used as the reference due to its symmetry and simplicity. The transformation $\mathcal{T}_{\mathbb{E}_{\mathbb{T}_\mathbf{O}} \to \mathbb{H}}(\mathbf{x})$ from tangent space $\mathbb{E}_{\mathbb{T}_\mathbf{O}}$ to hyperbolic space $\mathbb{H}$ is given by $\mathbf{x}\sinh(\sqrt{\rho}\|\mathbf{x}\|_{\mathbb{E}})/(\sqrt{\rho}\|\mathbf{x}\|_{\mathbb{E}})$.

2.2 Overview of HyperPath

As shown in Fig. 2, we transfer the latent knowledge from pathology Vision-Language Models (VLMs) and adapt it to hyperbolic space, and hierarchically aggregate features.

Specifically, the text encoder F_T extracts semantic features $f^T_{c,h}$ through $F_T([\tilde{p}_{c,h}, p_c])$ by combining concepts p_c with learnable prompts $\tilde{p}_{c,h}$. These features are adapted to tangent space via $Adaptor_T$ (a trainable two-layer MLP), producing $\tilde{f}^T_{c,h}$, and further transformed into hyperbolic embeddings $\mathbf{z}^T_{c,h}$ using $\mathcal{T}_{\mathbb{E}\to\mathbb{H}}$. For images, WSIs are divided into N_r regions (4096×4096), each split into N_p patches (512×512). Features f^I_p from F_I are mapped to $\tilde{f}^I_p \in \mathbb{E}_{\mathbb{T}_\mathbf{o}}$ via $Adaptor_I$ and embedded into $\mathbf{z}^I_p \in \mathbb{H}^k_\rho$, resulting in patch-level features of shape $N_r \times N_p \times D$.

Then, we design an aggregator AGG to integrate features $f^I_{h'} \in \mathbb{R}^{N_h \times N_{h'} \times D}$ from the **hierarchical subordinate level $\mathbf{h'}$** (patch/region) to generate representations $f^I_h \in \mathbb{R}^{N_h \times 1 \times D}$ for the current level h (region/slide). The process uses learnable weights $W_1 \in \mathbb{R}^{D/4 \times D}$, $W_2 \in \mathbb{R}^{D/4 \times 1}$, and is defined as:

$$f^I_h = \sum_{m=1}^{N_{h'}} \frac{\exp\left(W_2^\top \tanh(W_1 f^{I^\top}_{h',m})\right)}{\sum_{n=1}^{N_{h'}} \exp\left(W_2^\top \tanh(W_1 f^{I^\top}_{h',n})\right)} f^I_{h',m}. \tag{1}$$

After that, the aggregated features are subsequently mapped to hyperbolic space and optimized by the following Angular Modality Alignment Loss ($\mathcal{L}_{AMA}$) and Semantic Hierarchy Consistency Loss ($\mathcal{L}_{SHC}$). This process generates a more informative slide-level representation for the final prediction.

2.3 Angular Modality Alignment Loss

Aligning hierarchical visual and textual embeddings in hyperbolic space is crucial for cross-modal alignment, often achieved via contrastive learning methods like InfoNCE [21]. However, existing methods using geodesic distances [5] struggle with modality differences. Textual embeddings, which are more general, reside closer to the origin and entail the broad scope of concepts, while visual embeddings grow farther as granularity increases (e.g., slides→regions→patches). This leads to a mismatch: intra-modal geodesic distances differ in scale from inter-modal ones, disrupting proper alignment.

To address this, we propose Angular Modality Alignment Loss $\mathcal{L}_{ama}$, which leverages angular distance instead of geodesic distance. This provides a softer way to measure semantic similarity in hierarchical structures, enabling robust cross-modal alignment despite hierarchical differences across modalities. To be specific, we define **exterior angle** $\theta(\mathbf{u}, \mathbf{v})$ as:

$$\theta(\mathbf{u}, \mathbf{v}) = \pi - \angle \mathbf{Ouv} = \cos^{-1}\left(\frac{\mathbf{v}_t + \mathbf{u}_t \rho\langle \mathbf{u}, \mathbf{v}\rangle_\mathbb{H}}{\|\mathbf{u}_s\|_\mathbb{E}\sqrt{(\rho\langle \mathbf{u}, \mathbf{v}\rangle_\mathbb{H})^2 - 1}}\right). \tag{2}$$

Then, this exterior angle is used to compute the **angle distance** $\varphi(\mathbf{u},\mathbf{v}) = \theta(\mathbf{u},\mathbf{v})+\theta(\mathbf{v},\mathbf{u})-\pi$. After that, we define the **semantic similarity** as $\tilde{\varphi}(\mathbf{u},\mathbf{v}) = \varphi(\mathbf{v}^+,\mathbf{v}^-) - \varphi(\mathbf{u},\mathbf{v})$, where $\mathbf{v}^+$ and $\mathbf{v}^-$ are positive and negative hyperbolic embeddings for query $\mathbf{u}$. Through the defined semantic similarity, we aim to push the query away from negative samples while pulling it closer to positive ones, as shown in Fig. 2. To minimize angular distance for features of the same category, the alignment loss is formulated as:

$$\mathcal{L}_{ama}[\mathbf{u}|(\mathbf{v}^+,\mathbf{v}^-)] = -\log \frac{\exp(\tilde{\varphi}(\mathbf{u},\mathbf{v}^+)/\tau)}{\exp(\tilde{\varphi}(\mathbf{u},\mathbf{v}^+)/\tau) + \sum_{\mathbf{v}^-}\exp(|\tilde{\varphi}(\mathbf{u},\mathbf{v}^-)|/\tau)}, \quad (3)$$

where τ is the temperature. The absolute value of similarity between the query and negative samples penalizes cases where the query deviates from both negative and positive samples (Fig. 2), avoiding suboptimal convergence.

At each hierarchical level h, we apply a bidirectional alignment loss to visual and textual embeddings to mitigate bias. In the absence of specific labels for patches and regions, cosine similarity is computed using raw visual features and class semantic features extracted via CONCH. The loss is applied to the top-K most similar to their corresponding class semantics. Finally, the loss is $\mathcal{L}_{AMA} = \sum_h \Big(\mathcal{L}_{ama}[\mathbf{z}_h^I|(\mathbf{z}_{c^+,h}^T,\mathbf{z}_{c^-,h}^T)] + \mathcal{L}_{ama}[\mathbf{z}_{c^+,h}^T|(\mathbf{z}_h^I,\mathbf{z}_{c^-,h}^T)]\Big)$.

2.4 Semantic Hierarchy Consistency Loss

Beyond cross-modal alignment, capturing hierarchical semantics within and across modalities is crucial. Proper modeling ensures embeddings reflect structural dependencies and fine-grained details for coherent, interpretable representations. To achieve this, we introduce entailment cones in hyperbolic space to model partial order relationships and reinforce hierarchical consistency. The **half-aperture** is defined as $\phi(\mathbf{u}) = \sin^{-1}\left(2\alpha/(\sqrt{\rho}\|\mathbf{u}_s\|_{\mathbb{E}})\right)$, with $\alpha = 0.1$ to set boundary conditions near the origin [6]. Based on the definition, general concepts reside closer to the origin with wider apertures, while specific concepts are farther away with narrower apertures, reflecting their hierarchy in hyperbolic space.

To maintain semantic hierarchy consistency using entailment cones, we manage both intra-modal and inter-modal relations by explicitly addressing entailment and contradiction. For semantic entailment, $\mathbf{v}$ is entailed if it lies within the entailment cone of its hierarchical superordinate $\mathbf{u}$. For semantic contradiction where $\mathbf{u}$ shouldn't entail $\mathbf{v}$, we ensure $\mathbf{v}$ remains distant from the entailment cone boundary of $\mathbf{u}$, maintaining a clear separation between conflicting semantics and strengthening hierarchical consistency. These losses are formulated as:

$$\begin{cases} \mathcal{L}_{ent}(\mathbf{u},\mathbf{v}) = \exp\left(\theta\left(\mathbf{u},\mathbf{v}\right)/\phi\left(\mathbf{u}\right)-1\right)\cdot\max\left(\theta\left(\mathbf{u},\mathbf{v}\right)-\beta_{ent}\cdot\phi\left(\mathbf{u}\right),0\right) \\ \mathcal{L}_{con}(\mathbf{u},\mathbf{v}) = \exp\left(\phi\left(\mathbf{u}\right)/\theta\left(\mathbf{u},\mathbf{v}\right)-1\right)\cdot\max\left(\phi\left(\mathbf{u}\right)-\beta_{con}\cdot\theta\left(\mathbf{u},\mathbf{v}\right),0\right), \end{cases} \quad (4)$$

where the exponential function is employed to scale the penalty, β controls the margin, facilitating effective hierarchical semantic distinction. The final losses are defined as $\mathcal{L}_{ENT} = \sum_h \Big(\mathcal{L}_{ent}(\mathbf{z}_h^I,\mathbf{z}_{h'}^I) + \mathcal{L}_{ent}(\mathbf{z}_{c,h}^T,\mathbf{z}_{c,h'}^T) + \mathcal{L}_{ent}(\mathbf{z}_{c^+,h}^T,\mathbf{z}_h^I)\Big)$

and $\mathcal{L}_{CON} = \sum_h \left(\mathcal{L}_{con}(\mathbf{z}^T_{c^-,h}, \mathbf{z}^I_h) \right)$, which consist of intra- or inter-modal entailments and contradictions across hierarchical levels. Consequently, the semantic hierarchy consistency loss is given by $\mathcal{L}_{SHC} = \mathcal{L}_{ENT} + \mathcal{L}_{CON}$.

2.5 Slide-level Prediction

Slide-level classification is performed by leveraging semantic hierarchies in hyperbolic space, removing the dependency on additional linear classifiers. This is achieved using **geodesic** $d_G(\mathbf{u}, \mathbf{v}) = \sqrt{1/\rho} \cosh^{-1}(-\rho \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{H}})$, which measures distances between hyperbolic embeddings as curves. Let $d_G(\mathbf{z}^I_s, \mathbf{z}^T_{c_i,s})$ denote the geodesic distance between slide representation $\mathbf{z}^I_s$ and class-specific semantics $\mathbf{z}^T_{c_i,s}$ in hyperbolic space. The classification loss $\mathcal{L}_{CLS}$ is defined as:

$$\mathcal{L}_{CLS} = -\sum_{i=1}^{N_C} y_i \log \left(\frac{\exp(-d_G(\mathbf{z}^I_s, \mathbf{z}^T_{c_i,s}))}{\sum_{j=1}^{N_C} \exp(-d_G(\mathbf{z}^I_s, \mathbf{z}^T_{c_j,s}))} \right). \tag{5}$$

In summary, the overall loss is expressed as $\mathcal{L} = \mathcal{L}_{CLS} + \lambda_a \mathcal{L}_{AMA} + \lambda_s \mathcal{L}_{SHC}$, where λ_a and λ_s balance the contributions of losses.

3 Experiment

3.1 Experimental Settings

Datasets and Evaluation Metrics. We evaluated HyperPath's performance on four TCGA [30] tasks: breast cancer (BRCA) and non-small cell lung cancer (NSCLC) subtyping, HER2 [17] status prediction for breast cancer (BRCA HER2), and EGFR [4] mutation prediction for lung adenocarcinoma (LUAD EGFR). A nested splitting strategy was used: N_{outer} outer folds were generated based on Tissue Source Site codes, with one fold as in-domain (IND) and the rest as out-of-domain (OOD) from entirely different sites. Within each IND fold, N_{inner} inner Monte Carlo cross-validation splits were performed, resulting in $N_{outer} \times N_{inner}$ total folds. Performance was measured by mean and standard deviation of **AUC** ($\mathcal{A}$) and **F1 score** ($\mathcal{F}$) across all folds. We set $N_{outer} = 3$ and $N_{inner} = 5$ to ensure sufficient training samples due to limited EGFR data sites. For other tasks, $N_{outer} = 5$ and $N_{inner} = 3$ were used for a robust and comprehensive evaluation.

Implementation Details. Experiments were run on a single NVIDIA RTX 3090 GPU for 20 epochs using Adam optimizer [13] ($lr = 2 \times 10^{-4}$). Key hyperparameters were set as follows: $\tau = 0.05$, $\beta = 0.8$, $\lambda_a = 1$, and $\lambda_s = 10$.

3.2 Experimental Results

Comparison Results. As shown in Table 1, HyperPath was compared with state-of-the-art WSI analysis methods, including non-hierarchical (ABMIL [11],

Table 1. Performance evaluation for different tasks. (Best: **Bold**, Second: Underlined)

Method	BRCA TYPE				NSCLC TYPE			
	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$
ABMIL [11]	0.898 ±0.040	0.653 ±0.070	0.922 ±0.058	0.690 ±0.151	0.940 ±0.030	0.866 ±0.027	0.978 ±0.014	0.917 ±0.032
CLAM-SB [19]	0.911 ±0.030	0.695 ±0.061	**0.934** ±0.041	0.705 ±0.104	0.947 ±0.017	0.874 ±0.023	0.977 ±0.014	0.913 ±0.017
TransMIL [24]	0.914 ±0.027	0.667 ±0.046	0.923 ±0.047	0.706 ±0.116	0.938 ±0.019	0.857 ±0.026	0.979 ±0.020	0.926 ±0.038
DTFD-MIL [32]	0.904 ±0.019	0.634 ±0.031	0.916 ±0.056	0.676 ±0.130	0.947 ±0.025	0.871 ±0.031	0.977 ±0.016	0.907 ±0.028
ACMIL [33]	0.897 ±0.028	0.645 ±0.065	0.931 ±0.044	0.718 ±0.098	0.944 ±0.019	0.863 ±0.025	0.977 ±0.014	0.925 ±0.023
HIPT [2]	0.914 ±0.022	0.661 ±0.033	0.918 ±0.049	0.707 ±0.089	0.948 ±0.017	0.867 ±0.025	0.981 ±0.016	0.916 ±0.031
HIT [10]	0.922 ±0.015	0.679 ±0.059	0.920 ±0.055	0.705 ±0.136	0.938 ±0.021	0.864 ±0.030	0.982 ±0.020	0.933 ±0.034
HyperPath	**0.933** ±0.017	**0.696** ±0.060	**0.934** ±0.046	**0.750** ±0.086	**0.957** ±0.014	**0.883** ±0.017	**0.984** ±0.011	**0.938** ±0.026

Method	BRCA HER2				LUAD EGFR			
	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$
ABMIL [11]	0.660 ±0.049	0.184 ±0.070	0.681 ±0.157	0.210 ±0.200	0.611 ±0.039	0.308 ±0.088	0.595 ±0.107	0.315 ±0.107
CLAM-SB [19]	0.677 ±0.062	0.194 ±0.051	0.688 ±0.157	0.229 ±0.192	0.612 ±0.045	0.323 ±0.091	0.628 ±0.130	0.320 ±0.187
TransMIL [24]	0.734 ±0.049	0.192 ±0.085	0.700 ±0.160	0.152 ±0.163	0.626 ±0.038	0.294 ±0.104	0.619 ±0.110	0.306 ±0.164
DTFD-MIL [32]	0.712 ±0.060	0.232 ±0.052	0.719 ±0.138	0.218 ±0.136	0.573 ±0.057	0.290 ±0.123	0.626 ±0.113	0.310 ±0.146
ACMIL [33]	0.716 ±0.043	0.204 ±0.057	0.709 ±0.160	0.226 ±0.163	0.612 ±0.045	0.322 ±0.117	0.617 ±0.132	0.328 ±0.173
HIPT [2]	0.732 ±0.055	0.229 ±0.060	0.720 ±0.145	0.238 ±0.168	0.599 ±0.036	0.343 ±0.101	0.630 ±0.127	**0.358** ±0.149
HIT [10]	0.740 ±0.057	0.075 ±0.079	**0.740** ±0.144	0.093 ±0.178	**0.638** ±0.037	0.237 ±0.079	**0.647** ±0.111	0.256 ±0.224
HyperPath	**0.752** ±0.050	**0.260** ±0.086	0.732 ±0.157	**0.274** ±0.180	0.637 ±0.044	**0.378** ±0.093	0.638 ±0.107	0.343 ±0.120

CLAM-SB [19], TransMIL [24], DTFD-MIL [32], ACMIL [33]) and hierarchical approaches (HIT [10], HIPT [2]). All methods used CONCH [18] for feature extraction to ensure fair comparison. Notably, HyperPath achieves significant gains in both AUC and F1 Score across all tasks. In the OOD setting, it outperforms others by 1.9%–9.2% in AUC and 2.6%–8.8% in F1 Score (excluding HIT's outlier). Similarly, in the IND setting, it shows improvements of 0.7%–5.1% in AUC and 3.1%–12.2% in F1 Score. This consistent performance highlights its robustness with minimal variation between IND and OOD scenarios, except for a slight drop in $\mathcal{F}_{IND}$ on LUAD EGFR, likely due to small sample size and class imbalance. While HIT achieves high AUC in BRCA HER2 and LUAD EGFR, its low F1 score indicates a bias toward certain classes, reflecting overfitting and instability. In contrast, HyperPath delivers balanced and reliable results, excelling in both metrics and demonstrating its superiority across diverse tasks.

Table 2. Ablation study of HyperPath.

HyperPath		BRCA				NSCLC			
$\mathcal{L}_{AMA}$	$\mathcal{L}_{SHC}$	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$	$\mathcal{A}_{OOD}$	$\mathcal{F}_{OOD}$	$\mathcal{A}_{IND}$	$\mathcal{F}_{IND}$
		0.864 ±0.032	0.532 ±0.151	0.893 ±0.055	0.564 ±0.199	0.849 ±0.120	0.814 ±0.073	0.914 ±0.106	0.887 ±0.081
✓		0.925 ±0.016	0.634 ±0.107	0.928 ±0.046	0.656 ±0.150	0.947 ±0.021	0.832 ±0.114	0.975 ±0.024	0.889 ±0.121
	✓	0.539 ±0.105	0.156 ±0.141	0.558 ±0.123	0.153 ±0.139	0.499 ±0.099	0.302 ±0.326	0.507 ±0.156	0.331 ±0.331
✓	✓	**0.933** ±0.017	**0.696** ±0.060	**0.934** ±0.046	**0.750** ±0.086	**0.957** ±0.014	**0.883** ±0.017	**0.984** ±0.011	**0.938** ±0.026

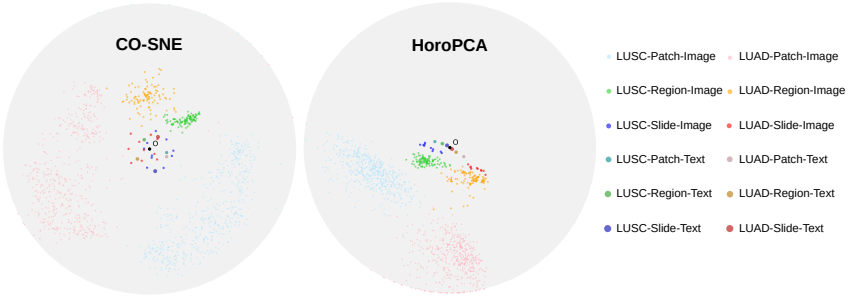

Fig. 3. The visualization of hyperbolic embeddings from different hierarchical levels. It is observed that the embeddings are well-structured in hyperbolic space.

Ablation Analysis. We conduct ablation studies to evaluate the effectiveness of $\mathcal{L}_{AMA}$ and $\mathcal{L}_{SHC}$ as shown in Tab. 2. Using $\mathcal{L}_{AMA}$ alone improves performance by aligning hyperbolic visual features with class semantics, while $\mathcal{L}_{SHC}$ alone degrades performance due to neglecting visual-textual alignment, leading to scattered feature distributions. Combining both losses achieves optimal results, as $\mathcal{L}_{AMA}$ ensures precise visual-semantic alignment and $\mathcal{L}_{SHC}$ enhances semantic hierarchies, promoting intra-semantic alignment, inter-semantic separation, and ultimately boosting classification performance through jointly learned multi-modal hyperbolic features with the consistent semantic hierarchy.

Hyperbolic Embedding Visualization. Fig. 3 visualizes hyperbolic embeddings using dimensionality reduction methods CO-SNE [7] and HoroPCA [1] designed for hyperbolic space. Using NSCLC subtyping task as an example, we display features across modalities, categories, and hierarchical levels. The visualization reveals a clear hierarchical structure: class semantic features cluster near the origin, surrounded by slide-, region-, and patch-level features in order. Distinct category boundaries confirm that our method effectively aligns and distributes multimodal features in hyperbolic space. Such hierarchical representations can further enhance WSI classification performance.

4 Conclusion

In this paper, we introduce HyperPath, a novel approach that leverages hyperbolic space to learn hierarchical representations of WSIs. HyperPath aggregates patch-level features from a pathology vision-language foundation model into region- and slide-level representations. Through angular modality alignment loss, semantically similar features are brought closer in hyperbolic space, while a semantic hierarchy consistency loss enhances inter- and intra-modality relationships, yielding meaningful hierarchies. Experiments demonstrate that HyperPath surpasses existing methods across multiple tasks, showing the potential of hyperbolic space as a powerful alternative for modeling complex WSIs.

Acknowledgments. This work was supported in part by the Research Grants Council of Hong Kong (27206123, C5055-24G, and T45-401/22-N), the Hong Kong Innovation and Technology Fund (ITS/274/22, and GHP/318/22GD), the National Natural Science Foundation of China (No. 62201483), and Guangdong Natural Science Fund (No. 2024A1515011875).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chami, I., Gu, A., Nguyen, D.P., Ré, C.: Horopca: Hyperbolic dimensionality reduction via horospherical projections. In: International Conference on Machine Learning. pp. 1419–1429. PMLR (2021)
2. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16144–16155 (2022)
3. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
4. da Cunha Santos, G., Shepherd, F.A., Tsao, M.S.: Egfr mutations and lung cancer. Annual Review of Pathology: Mechanisms of Disease **6**(1), 49–69 (2011)
5. 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)
6. Ganea, O., Bécigneul, G., Hofmann, T.: Hyperbolic entailment cones for learning hierarchical embeddings. In: International conference on machine learning. pp. 1646–1655. PMLR (2018)
7. Guo, Y., Guo, H., Yu, S.X.: Co-sne: Dimensionality reduction and visualization for hyperbolic data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21–30 (2022)
8. Guo, Z., Zhao, W., Wang, S., Yu, L.: Higt: Hierarchical interaction graph-transformer for whole slide image analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 755–764. Springer (2023)

9. Huang, Y., Zhao, W., Chen, Y., Fu, Y., Yu, L.: Free lunch in pathology foundation model: Task-specific model adaptation with concept-guided feature enhancement. Advances in Neural Information Processing Systems **37**, 79963–79995 (2025)
10. Huang, Y., Zhao, W., Wang, S., Fu, Y., Jiang, Y., Yu, L.: Conslide: Asynchronous hierarchical interaction transformer with breakup-reorganize rehearsal for continual whole slide image analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21349–21360 (2023)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
12. Khrulkov, V., Mirvakhabova, L., Ustinova, E., Oseledets, I., Lempitsky, V.: Hyperbolic image embeddings. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6418–6428 (2020)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Li, X., Li, C., Rahaman, M.M., Sun, H., Li, X., Wu, J., Yao, Y., Grzegorzek, M.: A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artificial Intelligence Review **55**(6), 4809–4878 (2022)
15. Liu, S., Chen, J., Pan, L., Ngo, C.W., Chua, T.S., Jiang, Y.G.: Hyperbolic visual embedding learning for zero-shot recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9273–9281 (2020)
16. Liu, Y., He, Z., Han, K.: Hyperbolic category discovery. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 9891–9900 (2025)
17. Loibl, S., Gianni, L.: Her2-positive breast cancer. The Lancet **389**(10087), 2415–2429 (2017)
18. Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. Nature Medicine **30**(3), 863–874 (2024)
19. 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)
20. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. Advances in neural information processing systems **10** (1997)
21. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
22. Pal, A., van Spengler, M., di Melendugno, G.M.D., Flaborea, A., Galasso, F., Mettes, P.: Compositional entailment learning for hyperbolic vision-language models. arXiv preprint arXiv:2410.06912 (2024)
23. 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)
24. 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)
25. Sinha, A., Zeng, S., Yamada, M., Zhao, H.: Learning structured representations with hyperbolic embeddings. Advances in Neural Information Processing Systems **37**, 91220–91259 (2025)
26. Srivastava, S., Wu, K.: Vision-language understanding in hyperbolic space (2024), `https://www.amazon.science/publications/vision-language-understanding-in-hyperbolic-space`

27. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Liu, S., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., et al.: Virchow: A million-slide digital pathology foundation model. arXiv preprint arXiv:2309.07778 (2023)
28. Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., Li, R., Tang, H., Wang, K., Li, Y., et al.: A pathology foundation model for cancer diagnosis and prognosis prediction. Nature **634**(8035), 970–978 (2024)
29. Wang, Z., Ramasinghe, S., Xu, C., Monteil, J., Bazzani, L., Ajanthan, T.: Learning visual hierarchies with hyperbolic embeddings. arXiv preprint arXiv:2411.17490 (2024)
30. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature genetics **45**(10), 1113–1120 (2013)
31. Yang, M., Zhou, M., Ying, R., Chen, Y., King, I.: Hyperbolic representation learning: Revisiting and advancing. In: International Conference on Machine Learning. pp. 39639–39659. PMLR (2023)
32. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18802–18812 (2022)
33. Zhang, Y., Li, H., Sun, Y., Zheng, S., Zhu, C., Yang, L.: Attention-challenging multiple instance learning for whole slide image classification. In: European Conference on Computer Vision. pp. 125–143. Springer (2024)