MaSH: Maximal Separating Poincaré Hyperplanes for Hierarchical Imbalanced Learning

Bo Xiong
Stanford University
Palo Alto, USA
xiongbo@stanford.edu

Jiaying Lu
Emory University
Atlanta, USA
jiaying.lu@emory.edu

Yuqicheng Zhu
Bosch Center for AI
Stuttgart, Germany
yuqicheng.zhu@ki.uni-stuttgart.de

Carl Yang
Emory University
Atlanta, USA
j.carlyang@emory.edu

ABSTRACT

Hyperbolic classifiers, which typically view hyperbolic hyperplanes as decision boundaries, are generalizations of Euclidean classifiers in hyperbolic space suitable for modeling hierarchical data. In this paper, we present **Ma**ximal **S**eparating Poincaré **H**yperplane (**MaSH**), a hyperbolic geometric inductive bias that enhances the generalization capability of hyperbolic classifiers, especially on the class-imbalanced settings. MaSH encourages 1) the *equiangularity* of the ideal points of the Poincaré hyperplanes of all classes; and 2) the *equiradiality* of these Poincaré hyperplanes. The two properties jointly encourage the *maximal separation bias* for hyperbolic classifiers. We perform experiments on imbalanced/long-tailed classification and the results show consistent improvements.

CCS CONCEPTS

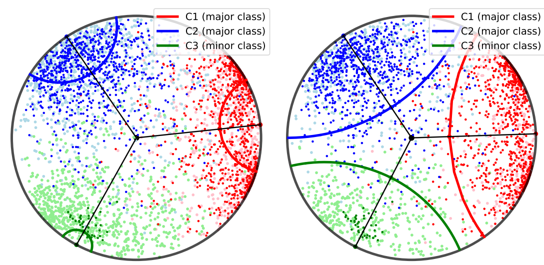
• **Computing methodologies** → **Learning linear models**.

KEYWORDS

Hyperbolic classification, imbalanced classification, long-tailed learning, hierarchical classification

ACM Reference Format:

Bo Xiong, Jiaying Lu, Yuqicheng Zhu, and Carl Yang. 2025. MaSH: Maximal Separating Poincaré Hyperplanes for Hierarchical Imbalanced Learning. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3701716.3715505



(a) HMLR (Acc: 75.7) **(b) MaSH (Acc: 91.2)**

Figure 1: A synthetic example on imbalanced classification in a 2D hyperbolic space, where $C1$ and $C2$ are major classes and $C3$ is a minor class with only 5% training examples of the major classes. Training and testing examples of each class are colored by dark and light colors, respectively. (a) HMLR does not perform well on the minor class; (2) The proposed MaSH improves the accuracy by 20%+.

1 INTRODUCTION

Learning representations in hyperbolic space has gained popularity due to its advantageous properties for encoding hierarchical data [23, 25, 29, 30, 34]. The ability to encode hierarchies stems from the fact that the volume of hyperbolic space grows exponentially with an increase in radius, mirroring the discrete property of trees where the number of nodes grows exponentially with depth. Hyperbolic embedding projects data into hyperbolic space, and it has been successfully applied to represent various forms of hierarchical data, including images [12], texts [23], networks [2], and knowledge graphs [29].

By embedding Euclidean space features into hyperbolic space, conventional machine learning algorithms can be extended to operate within hyperbolic geometry through the definition of equivalent vector operations. Hyperbolic classifiers [3, 5, 7, 8, 24, 25] generalize their Euclidean counterparts by learning decision boundaries in hyperbolic space, and have demonstrated superior performance on datasets characterized by hierarchical semantics. As a counterpart to multinomial logistic regression (MLR) in Euclidean space, Hyperbolic MLR (HMLR) has served as a standard classifier layer in various hyperbolic learning architectures, including hyperbolic image classifiers and hyperbolic graph networks [19, 30, 34, 35]. In HMLR [8], classification logits are formulated as distances from instance embeddings to margin hyperplanes in hyperbolic space.

Concerning the generalization of classifiers in the imbalanced settings, one might introduce an inductive bias that describes a better assumptions about the target classifiers independent of the training data. One of the most prominent inductive bias is *maximal class separation* bias [11]. Given many possible classifiers, the maximal class separation bias is to select the one that represents

WWW Companion '25, April 28-May 2, 2025, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1331-6/25/04
https://doi.org/10.1145/3701716.3715505

the maximal pair-wise separation of class decision boundaries. To achieve this, a wide range of works have investigated optimization strategies to explicitly enforce classes away from each other. A common approach is to introduce additional losses [15–18] which incorporate a notion of class margins. However, the generalization ability of classifiers in hyperbolic space has not yet been explored.

This paper explores the imbalanced (or long-tailed) classification problem in hyperbolic space and presents **Ma**ximal **S**eparating Poincaré **H**yperplane (**MaSH**), a geometric inductive bias that enforces the class hyperplanes in hyperbolic space to be maximally separated. Fig. 1 shows a comparison of MaSH and HMLR on imbalanced classification. MaSH is inspired by the maximal class separation bias but is fully designed in hyperbolic geometry. Different from Euclidean geometry, each hyperplane in hyperbolic space is a curved subspace and it can be uniquely specified by two features: 1) the direction of the hyperplane which specifies where the hyperplane points to and each direction can be uniquely specified by an ideal point in the boundary of the Poincaré ball; and 2) the radius of the hyperplane which specifies the volume that the convex regions of the hyperplane encompass, where the radius can be calculated as the distance from the hyperplane to the origin. Our proposed MaSH is designed to encourage the seperation/uniformity of both features. In particular, MaSH enforces 1) the **equiangularity** (i.e., equal pair-wise angle) of the directions of the Poincaré hyperplanes of all classes; and 2) the **equiradiality** (i.e., equal radius) of these Poincaré hyperplanes. The satisfaction of these two properties jointly describe a maximal separation bias for hyperbolic classifiers. We propose two regularization terms to encourage equiangularity and equiradiality, respectively. We conduct experiments on imbalanced/long-tailed classification. The experimental results demonstrate that MaSH outperforms HMLR, especially in the imbalanced or long-tailed settings. Our code for reproducing the results will be open upon acceptance.

2 RELATED WORK

Classification in hyperbolic space. Hyperbolic geometry provides an alternative space for representing data whose samples or labels exhibit a hierarchical structure [1, 2, 2, 8, 12, 14, 21, 22, 26–28, 31, 32]. Euclidean classification layers have also been generalized into hyperbolic space [5, 7]. HMLR [8] generalizes multinomial logistic regression (MLR) by formulating logits as distances from instance embedding to the margin hyperplane in hyperbolic space, which is defined as the intersection of the hyperboloid model and a hyperplane in the ambient space, a.k.a. geodesic hyperplane. Hyperbolic SVM [3, 5, 24] considers the maximum margin learning bias and can be viewed as a hyperbolic formulation of support vector machine (SVM) classifiers. HoroSVM [7], which is also a formulation of SVM in hyperbolic space, models classification hyperplanes as horospheres instead of geodesic hyperplanes. Following these works, more complicated classifiers such as hyperbolic decision tree and random forest [4, 6] have also been proposed. However, none of the existing work has considered the imbalanced setting of hyperbolic classifiers, which is a more realistic setting in our real-world applications.

3 PRELIMINARIES

Poincaré ball model The Poincaré ball $\left(\mathbb{D}^n, g^{\mathbb{D}}\right)$ is one of the models of hyperbolic geometry. The Poincaré ball is defined as an open n-ball $\mathbb{D}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\}$ equipped with a Riemannian metric $g_{\mathbf{x}}^{\mathbb{D}} = \lambda_{\mathbf{x}}^2 g^E$, where $\lambda_{\mathbf{x}} = \frac{2}{1-\|\mathbf{x}\|^2}$, $g^E = \mathbf{I}_n$ is the Euclidean metric tensor, $\lambda_{\mathbf{x}}$ is the *conformal factor*, and $\|\cdot\|^2$ denotes the L^2 norm in Euclidean space. The distance between two points $\mathbf{x}, \mathbf{y} \in \mathbb{D}^n$ can be defined by $d_{\mathbb{D}}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left(1 + 2\frac{\|\mathbf{x}-\mathbf{y}\|^2}{(1-\|\mathbf{x}\|^2)(1-\|\mathbf{y}\|^2)}\right)$.

Hyperbolic MLR (HMLR) generalizes Euclidean MLR to the hyperbolic space by viewing Poincaré hyperplane as linear decision boundaries. A Poincaré hyperplane is defined as the intersection of a Euclidean subspace and the Poincaré ball plus all linear subspaces going through the origin. For the former cases, a Poincaré hyperplane can be uniquely defined by its center point that has a minimal distance to the origin.

Definition 1 (Poincaré hyperplane). *Given a (center) point $\mathbf{c} \in \mathbb{D}^n$ where $\mathbf{c} \neq \mathbf{0}$, the Poincaré hyperplane is defined as*

$$H^{\mathbf{c}} = \left\{\mathbf{p} \in \mathbb{D}^n : g^{\mathbb{D}}\left(\log_{\mathbf{c}}(\mathbf{p}), \vec{\mathbf{c}}\right) = 0\right\}, \tag{1}$$

where $\mathbf{c}$ is the center point and $\vec{\mathbf{c}} \in T_{\mathbf{c}}\mathbb{D}^n$.

Given the definition of Poincaré hyperplane, HMLR is defined as:

Definition 2 (HMLR). *Let $\mathcal{X} \subseteq \mathbb{R}^n$ denote an n-dimensional instance space, $C \doteq \{1, \ldots, K\}$ denote a finite set of possible classes where K is the number of classes. Given a set of N training examples $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid 1 \leq i \leq N, \mathbf{x}_i \in \mathcal{X}, y_i \in C\}$. HMLR seeks to learn a transformation function $f_{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{D}^n$ that maps inputs to a Poincaré ball and a set of linear classifiers that correctly classify the training examples. This is typically achieved by training a model with the optimization objective*

$$\mathcal{L}_{\text{HMLR}} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\left(\left[v_i^1, \ldots, v_i^K\right], y_i\right), \text{ s.t. } \mathbf{u}_i = f_{\boldsymbol{\theta}}(\mathbf{x}_i), \tag{2}$$

where $f_{\boldsymbol{\theta}}$ can be viewed as a feature extractor implemented by a hyperbolic neural network with trainable parameters $\boldsymbol{\theta}$. The output $\boldsymbol{u}_i$ of the feature extractor is referred to as the feature of $\boldsymbol{x}_i$. The loss function $\mathcal{L}(\cdot,\cdot)$ is typically defined as the cross-entropy loss. The vector $[v_{i1}, \ldots, v_{iK}]$ is often referred to the logit vector for data sample $\boldsymbol{x}_i$. The logit *function is calculated as $v_i^k = \text{logit}(\mathbf{u}_i, H_k)$, which is the distance from the sample feature $\mathbf{u}_i$ to a Poincaré hyperplane H_k. The distance has the closed form $d(\mathbf{u}, H^{\mathbf{c}}) = \sinh^{-1}\left(\frac{2|\langle(-\mathbf{c})\oplus\mathbf{u}, \mathbf{c}\rangle|}{(1-\|(-\mathbf{c})\oplus\mathbf{u}\|^2)\|\mathbf{c}\|}\right)$.*

4 MASH: MAXIMAL SEPARATING POINCARÉ HYPERPLANE

The geometric intuition of MaSH is based on two intuitions: 1) maximally separating the directions of classifiers, which is equivalent to enforcing hyperspherical uniformity of the ideal points of the Poincaré hyperplanes. 2) encouraging fairness of these classifiers by imposing equiradiality (equal radii) of the Poincaré hyperplanes.

4.1 The Geometric Structure of MaSH

We first consider the structure of simplex equiangular tight frame (ETF) [36], in which the vectors have equal pair-wise angles.

Definition 3 (Simplex Equiangular Tight Frame). *A collection of vectors $\{\zeta_i\}_{i=1}^K$ where $\zeta_i \in \mathbb{R}^d, d \geq K$, is said to be simplex equiangular tight frame if:*

$$\mathbf{Z} = \sqrt{\frac{K}{K-1}}\mathbf{U}\left(\mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^T\right). \tag{3}$$

where $\mathbf{Z} = [\zeta_1, \cdots, \zeta_K] \in \mathbb{R}^{d\times K}, \mathbf{U} \in \mathbb{R}^{d\times K}$ satisfies $\mathbf{U}^T\mathbf{U} = \mathbf{I}_K, \mathbf{I}_K$ is the identity matrix, and $\mathbf{1}_K$ is vector with all elements being one

All vectors in a simplex ETF have an equal ℓ_2 norm and the same pair-wise angle, i.e., $\zeta_i^T\zeta_j = \frac{K}{K-1}\delta_{i,j} - \frac{1}{K-1}, \forall i,j \in [1,K]$, where $\delta_{i,j}$ equals to 1 when $i=j$ and 0 otherwise. The pair-wise angle $-\frac{1}{K-1}$ is the maximal equiangular separation of K vectors in $\mathbb{R}^d$. Hence, simplex ETF can be viewed as the maximal separation structure for linear classification. However, simplex ETF only exists when $d \geq K$, which is not possible when the number of classes is very large. This paper considers a more general cases $K \geq d$ and studies Grassmannan frame, which is defined as:

Definition 4 (Grassmannan Frame). *Given a unit norm frame defined as a sequence of K vectors $\{\zeta_i\}_{i=1}^K$ whose norms are all equal to 1, the frame is said to be a Grassmannan frame iff the frame is the solution of minimal maximal frame correlation*

$$\min\{\mathcal{M}\left(\{\zeta_i\}_{i=1}^K\right) = \max_{i,j,i\neq j}\left\{\left|\left\langle\zeta_i,\zeta_j\right\rangle\right|\right\}\}, \tag{4}$$

where the minimum is taken over all unit norm frames in $\mathbb{R}^d$, $\langle\cdot,\cdot\rangle$ denotes the inner product between vectors.

Essentially, Grassmannan frame is a relaxed version of simplex equiangular tight frame (ETF) [36] in which all vectors have equal pair-wise angles and it satisfies two important properties: 1) it has minimized maximal correlation which corresponds to hyperspherical uniformity; and 2) there exists at least one solution for any vector number K and dimension d, even when $K \geq d$.

Lemma 1 (Existence of Grassmannian frames [9]). *For any given vector number K and dimension d where $K \geq d$, there exists a Grassmannian frame denoted by* $\mathrm{GF}(K,d)$.

Given a Grassmannan frame, a MaSH can be defined as:

Definition 5 (MaSH). *A set of features $\{\boldsymbol{u}_i\}_{i=1}^N \subseteq \mathbb{D}^d$ in hyperbolic space is said to have a Maximal-Separating-Poincaré-Hyperplane (MaSH) arrangement with respect to a set of classes $\{y_i\}_{i=1}^K \subseteq [K]$ if and only if there exist a collection of Poincaré hyperplanes $\left\{H_{\mathbf{c}_k}\right\}_{k=1}^K$ parametrized by the center points $\{\mathbf{c}_k\}_{k=1}^K$, such that 1) these hyperplanes have equal radius; 2) the corresponding ideal points set $\boldsymbol{p} = \left\{\boldsymbol{p}_k = \frac{\boldsymbol{c}_k}{\|\boldsymbol{c}_k\|}\right\}_{k=1}^K$ of the center points form a Grassmannan frame; and 3) these Poincaré hyperplanes can correctly classify the input features, i.e., for all i, k that satisfies $y_i = k$, it has*

$$\arg\min_{k=1,2,\cdots,K} d(\boldsymbol{u}_i, H_{\mathbf{c}_k}) = k. \tag{5}$$

The MaSH arrangement has a simple geometric description. In short, it requires that features associated with each class lie inside the convex hull formed by a hyperplane while being outside of the convex hulls formed by other hyperplanes, and the hyperplanes have the properties that they have the same radius and their ideal points form a Grassmannan frame, hence different classes are sufficiently separated. Under the unconstrained feature models [36], which assume that the encoder $f(\cdot)$ can produce any set of features given any set of inputs, we have the following guaranteed existence of MaSH arrangement given any inputs.

Lemma 2 (Existence of MaSH arrangement). *For any given vector number C and dimension d where $K \geq d$, there must exist a MaSH arrangement under the unconstrained feature model.*

Proof. Given a vector number C and the dimension d where $C \geq d$, according to Lemma 1, there must exist a Grassmannian frame $\mathrm{GF}(C,d)$, such that a MaSH can be constructed by 1) setting the ideal points of the hyperplane centers as the vector sets of $\mathrm{GF}(C,d)$; and 2) requiring that these hyperplane centers having equal radius. □

4.2 Learning MaSH Structure for Classification

Although MaSH arrangements exist for any K and d where $K \geq d$ unconstrained feature model, constructing one from real-world datasets, which is as known as the Tammes problem [20], is a challenging problem. We hence consider gradient descent learning as an approximation. We design two regularization terms, equiangularity and equiradiality, which explicitly encourage the satisfaction of the MaSH structure.

Equiangularity encourages the maximal separation of the ideal points of hyperplanes. This can be achieved by minimizing the maximal cosine similarity between pairs of ideal points.

$$\mathbf{p}^* = \arg\min_{\mathbf{p}'\in\mathbb{P}}\left(\max_{(k,j,k\neq j)\in[K]}\cos\left(\mathbf{p}'_k,\mathbf{p}'_j\right)\right), \tag{6}$$

where $\mathbb{P}$ denote the solution space. However, optimizing this min-max problem is inefficient as it requires computing all pairwise cosine similarities and only the largest pair can be updated at each time step. Hence, we consider the following proxy loss:

$$\mathcal{L}_{\text{equiangularity}} = \frac{1}{K}\sum_{i=1}^K \max_{j\in[K]}\left(\mathbf{p}\mathbf{p}^T - 2\mathbf{I}\right)_{ij}, \tag{7}$$

where $\mathbf{p} \in \mathbb{R}^{K\times D}$ denotes the current set of ideal points, $\mathbf{I}$ denotes the identity matrix. The loss function minimizes the smallest cosine similarity for each pair of ideal points and can be optimized quickly through matrix computation.

Equiradiality encourages the volumes of convex hulls formed by classification hyperplanes to be as equal as possible. This can be achieved by minimizing the variances of the radii of hyperplanes,

$$\mathcal{L}_{\text{equiradiality}} = \frac{1}{K}\sum_{i=1}^K(\mathbf{p}_i - \hat{\mathbf{p}}), \tag{8}$$

where $\hat{\mathbf{p}}$ denote the average of ideal point set $\mathbf{p}$.

MaSH classifier. The loss for MaSH classifier can be defined as a weighted sum of the HMLR loss and the two regularization terms.

$$\mathcal{L}_{\text{MaSH}} = \mathcal{L}_{\text{HMLR}} + \lambda_1\mathcal{L}_{\text{equiangularity}} + +\lambda_2\mathcal{L}_{\text{equiradiality}}, \tag{9}$$

where λ_1 and λ_2 denote the corresponding regularization weights. Lemma 3 shows the soundness of the regularization terms.

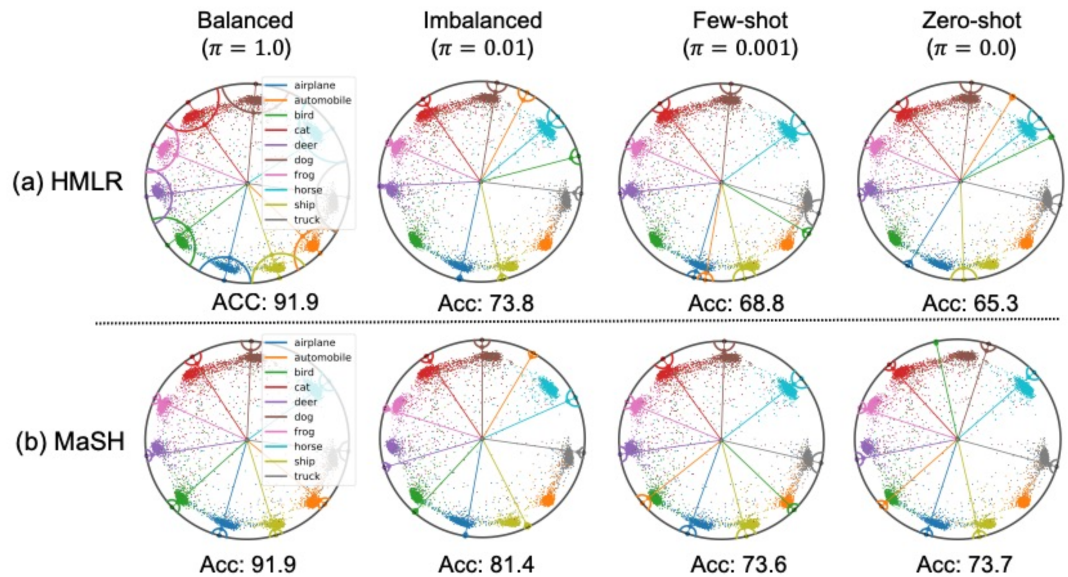


Figure 2: A visualization of the learned hyperbolic hyperplanes of HMLR and MaSH. For imbalanced, few-shot, and zero-shot learning settings, *airplane*, *automobile*, and *bird* are chosen as three minor classes.

Table 1: Imbalanced/long-tailed classification accuracy (%) with ResNet encoder on CIFAR-10. The numbers in the second row denote the imbalance ratio calculated as $\pi = \frac{n_{min}}{n_{max}}$ where n_{min} and n_{max} are the minimal and maximal numbers of training samples in all classes, respectively. Results are averaged by 5 repeated experiments with different seeds. Mixup is a regularization term used to improve the generalization of neural network architectures by adversarial training samples.

Methods	without Mixup					with Mixup				
	0.005	0.01	0.02	0.1	balanced	0.005	0.01	0.02	0.1	balanced
MLR	66.1	71.0	77.1	87.4	**93.4**	67.3	72.8	78.6	87.7	**93.6**
HMLR	66.5	72.2	77.7	87.5	93.1	68.4	74.6	79.6	87.9	93.3
MaSH (ours)	**67.2**	**72.9**	**78.4**	**87.7**	93.1	**68.8**	**75.0**	**79.9**	**88.1**	93.4
Δ(MaSH - HMLR)	1.1%	1.0%	0.9%	0.2%	0.0%	0.5%	0.5%	0.4%	0.2%	0.1%

Lemma 3 (soundness). *The MaSH classifier satisfies the MaSH arrangement if and only if* $\mathcal{L}_{\text{equiangularity}} = 0$ *and* $\mathcal{L}_{\text{equiradiality}} = 0$.

Proof. If $\mathcal{L}_{\text{equiangularity}} = 0$, then the ideal points of these hyperplanes form a Grassmannan frame. If $\mathcal{L}_{\text{equiradiality}} = 0$, then these hyperplanes have equal radius. Based on the definition of MaSH, these hyperplanes form a MaSH arrangement under the unconstrained feature model. Also, if the MaSH classifier satisfies the MaSH arrangement, based on the definition of the two regulization terms, we also have $\mathcal{L}_{\text{equiangularity}} = 0$ and $\mathcal{L}_{\text{equiradiality}} = 0$. □

5 EXPERIMENTS

We evaluate our method on multi-class imbalanced classification on CIFAR-10 [13]. We first test our method on CIFAR-10 dataset with pre-trained low-dimensional features in hyperbolic space, and train the classifiers in a 2-dimensional hyperbolic space. Fig. 2 shows a comparison of the learned hyperbolic hyperplanes of HMLR [8] and MaSH. First, in the balanced case, HMLR is sufficiently effective in classification and MaSH achieves the same performance. However, the learned hyperplanes of MaSH are clearly different from the ones of HMLR. In particular, the hyperplanes of MaSH are more far away from the origin and their volumes have small variances. This makes sense as our loss term encourages equal norms. Secondly, MaSH significantly improves HMLR on the imbalanced settings and even on few-shot/zero-shot settings. In particular, on the imbalanced setting ($\pi = 0.01$), HMLR fails in classifying two minor classes – *automobile* and *bird*, while MaSH only fails in classifying *automobile*. On the few-shot and zero-shot settings, both HMLR and MaSH can only correctly classify one minor class *airplane*, but MaSH's hyperplanes distribute more uniformly in the hyperbolic space and achieve better accuracy.

Table 2: The impact of the regularization terms in MaSH.

	0.005	0.01	0.02	0.1	balanced
HMLR	66.1	71.0	77.1	87.4	**93.4**
MaSH (w/o EquiAngle)	66.8	71.5	77.8	87.6	**93.4**
MaSH (w/o EquiRadius)	66.7	71.8	77.7	87.5	93.3
MaSH	**67.2**	**72.9**	**78.4**	**87.7**	93.1

Next, we test our model on CIFAR-10 dataset by training the ResNet [10] encoder and map the feature into a hyperbolic space with a hyperbolic linear layer. For ablation analysis, we set various imbalance ratios and conduct two experiments, one with Mixup [33] regularization and one without Mixup. As shown in Table 1, it is clear that HMLR outperforms Euclidean MLR in most of the imbalanced cases, demonstrating the advantages of hyperbolic classifiers.

It is also clear that the proposed method MaSH consistently improves over HMLR on imbalanced cases while achieving similar results on balanced cases, which is consistent with our major claim that MaSH generalizes better. We also find that the regularization Mixup does improve the generalization, but MaSH further improves it, which means that the benefits provided by Mixup (adversarial sampling) and MaSH can complement each other. Another interesting finding is that there is a positive correlation between performance gain and the imbalance ratio, which further demonstrates the advantages of MaSH in imbalanced classification settings.

Influence of regularization terms. Table 2 shows the performance of MaSH by removing one of the two regularization terms on the without-Mixup settings. It shows that both terms result in some improvements on the classification performance on the imbalanced settings. The combination of the two terms further boost the performance, demonstrating the unique benefit of both terms.

6 CONCLUSION

This paper introduces a hyperbolic geometric inductive bias suitable for hierarchical imbalanced classification. The inductive bias is inspired by the maximum separation bias and is fully defined in hyperbolic geometry. We propose MaSH, a hyperbolic classification approach with two regularization terms that encourage this geometric inductive bias. Our results showcase the advantages of the proposed MaSH on imbalanced or long-tailed classification. The MaSH framework can enhance AI-driven diagnostics in RADx projects by improving classification performance on imbalanced medical datasets, such as distinguishing rare COVID-19 positive cases from the majority of negative samples.

ACKNOWLEDGMENT

This research was partially supported by the US National Science Foundation under Award Number 2319449 and Award Number 2312502, as well as the US National Institute of Diabetes and Digestive and Kidney Diseases of the US National Institutes of Health under Award Number K25DK135913.

REFERENCES

[1] Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-Dimensional Hyperbolic Knowledge Graph Embeddings. In *ACL*. Association for Computational Linguistics, 6901–6914.

[2] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic Graph Convolutional Neural Networks. In *NeurIPS*. 4869–4880.

[3] Eli Chien, Chao Pan, Puoya Tabaghi, and Olgica Milenkovic. 2021. Highly Scalable and Provably Accurate Classification in Poincaré Balls. In *ICDM*. IEEE, 61–70.

[4] Philippe Chlenski, Ethan Turok, Antonio Khalil Moretti, and Itsik Pe'er. 2023. Fast hyperboloid decision tree algorithms. *CoRR* abs/2310.13841 (2023).

[5] Hyunghoon Cho, Benjamin Demeo, Jian Peng, and Bonnie Berger. 2019. Large-Margin Classification in Hyperbolic Space. In *AISTATS (Proceedings of Machine Learning Research, Vol. 89)*. PMLR, 1832–1840.

[6] Lars Doorenbos, Pablo Márquez-Neila, Raphael Sznitman, and Pascal Mettes. 2023. Hyperbolic Random Forests. *CoRR* abs/2308.13279 (2023).

[7] Xiran Fan, Chun-Hao Yang, and Baba C. Vemuri. 2023. Horospherical Decision Boundaries for Large Margin Classification in Hyperbolic Space. In *NeurIPS*.

[8] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic Neural Networks. In *NeurIPS*. 5350–5360.

[9] Arvind Ganesh. 2007. A brief introduction to grassmannian frames. (2007).

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

[11] Tejaswi Kasarla, Gertjan J. Burghouts, Max van Spengler, Elise van der Pol, Rita Cucchiara, and Pascal Mettes. 2022. Maximum Class Separation as Inductive Bias in One Matrix. In *NeurIPS*.

[12] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan V. Oseledets, and Victor S. Lempitsky. 2020. Hyperbolic Image Embeddings. In *CVPR*. Computer Vision Foundation / IEEE, 6417–6427.

[13] Alex Krizhevsky. 2009. Learning multiple layers of features from tiny images. (2009).

[14] Eric Wonhee Lee, Bo Xiong, Carl Yang, and Joyce C. Ho. 2024. HypMix: Hyperbolic Representation Learning for Graphs with Mixed Hierarchical and Non-hierarchical Structures. In *CIKM*. ACM, 3852–3856.

[15] Xiaoxu Li, Dongliang Chang, Zhanyu Ma, Zheng-Hua Tan, Jing-Hao Xue, Jie Cao, Jingyi Yu, and Jun Guo. 2020. OSLNet: Deep Small-Sample Classification With an Orthogonal Softmax Layer. *IEEE Trans. Image Process.* 29 (2020), 6482–6495.

[16] Rongmei Lin, Weiyang Liu, Zhen Liu, Chen Feng, Zhiding Yu, James M. Rehg, Li Xiong, and Le Song. 2020. Regularizing Neural Networks via Minimizing Hyperspherical Energy. In *CVPR*. Computer Vision Foundation / IEEE, 6916–6925.

[17] Weiyang Liu, Rongmei Lin, Zhen Liu, Li Xiong, Bernhard Schölkopf, and Adrian Weller. 2021. Learning with Hyperspherical Uniformity. In *AISTATS (Proceedings of Machine Learning Research, Vol. 130)*. PMLR, 1180–1188.

[18] Weiyang Liu, Yan-Ming Zhang, Xingguo Li, Zhen Liu, Bo Dai, Tuo Zhao, and Le Song. 2017. Deep Hyperspherical Learning. In *NIPS*. 3950–3960.

[19] Federico López and Michael Strube. 2020. A Fully Hyperbolic Neural Model for Hierarchical Multi-class Classification. In *EMNLP (Findings) (Findings of ACL, Vol. EMNLP 2020)*. Association for Computational Linguistics, 460–475.

[20] Edith Mooers. 1994. Tammes's problem. *University of Vermont* (1994).

[21] Mojtaba Nayyeri, Bo Xiong, Majid Mohammadi, Mst. Mahfuja Akter, Mirza Mohtashim Alam, Jens Lehmann, and Steffen Staab. 2023. Knowledge Graph Embeddings using Neural Ito Process: From Multiple Walks to Stochastic Trajectories. In *ACL (Findings)*. Association for Computational Linguistics, 7165–7179.

[22] Yanchao Tan, Hang Lv, Zihao Zhou, Wenzhong Guo, Bo Xiong, Weiming Liu, Chaochao Chen, Shiping Wang, and Carl Yang. 2024. Logical Relation Modeling and Mining in Hyperbolic Space for Recommendation. *The 40th IEEE International Conference on Data Engineering* (2024).

[23] Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. 2019. Poincare Glove: Hyperbolic Word Embeddings. In *ICLR (Poster)*. OpenReview.net.

[24] Melanie Weber, Manzil Zaheer, Ankit Singh Rawat, Aditya Krishna Menon, and Sanjiv Kumar. 2020. Robust large-margin learning in hyperbolic space. In *NeurIPS*.

[25] Bo Xiong, Michael Cochez, Mojtaba Nayyeri, and Steffen Staab. 2022. Hyperbolic Embedding Inference for Structured Multi-Label Prediction. In *NeurIPS*.

[26] Bo Xiong, Mojtaba Nayyeri, Daniel Daza, and Michael Cochez. 2023. Reasoning beyond Triples: Recent Advances in Knowledge Graph Embeddings. In *CIKM*. ACM, 5228–5231.

[27] Bo Xiong, Mojtaba Nayyeri, Ming Jin, Yunjie He, Michael Cochez, Shirui Pan, and Steffen Staab. 2023. Geometric Relational Embeddings: A Survey. *CoRR* abs/2304.11949 (2023).

[28] Bo Xiong, Mojtaba Nayyeri, Linhao Luo, Zihao Wang, Shirui Pan, and Steffen Staab. 2024. NestE: Modeling Nested Relational Structures for Knowledge Graph Reasoning. *AAAI* abs/2312.09219 (2024).

[29] Bo Xiong, Shichao Zhu, Mojtaba Nayyeri, Chengjin Xu, Shirui Pan, Chuan Zhou, and Steffen Staab. 2022. Ultrahyperbolic Knowledge Graph Embeddings. In *KDD*. ACM, 2130–2139.

[30] Bo Xiong, Shichao Zhu, Nico Potyka, Shirui Pan, Chuan Zhou, and Steffen Staab. 2022. Pseudo-Riemannian Graph Convolutional Networks. In *NeurIPS*.

[31] Menglin Yang, Aosong Feng, Bo Xiong, Jiahong Liu, Irwin King, and Rex Ying. 2024. Enhancing llm complex reasoning capability through hyperbolic geometry. In *ICML 2024 Workshop on LLMs and Cognition*.

[32] Menglin Yang, Aosong Feng, Bo Xiong, Jihong Liu, Irwin King, and Rex Ying. 2024. Hyperbolic Fine-tuning for Large Language Models. *CoRR* abs/2410.04010 (2024).

[33] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2018. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*.

[34] Min Zhou, Menglin Yang, Bo Xiong, Hui Xiong, and Irwin King. 2023. Hyperbolic Graph Neural Networks: A Tutorial on Methods and Applications. In *KDD*. ACM, 5843–5844.

[35] Shichao Zhu, Shirui Pan, Chuan Zhou, Jia Wu, Yanan Cao, and Bin Wang. 2020. Graph geometry interaction learning. *Advances in Neural Information Processing Systems* 33 (2020), 7548–7558.

[36] Zhihui Zhu, Tianyu Ding, Jinxin Zhou, Xiao Li, Chong You, Jeremias Sulam, and Qing Qu. 2021. A Geometric Analysis of Neural Collapse with Unconstrained Features. In *NeurIPS*. 29820–29834.