HYPERANFIS: ENHANCING RULE REPRESENTATION AND INTERPRETABILITY IN ADAPTIVE NEURO-FUZZY SYSTEMS VIA HYPERBOLIC GEOMETRY

Anonymous authors
Paper under double-blind review

ABSTRACT

The adaptive neuro-fuzzy inference system (ANFIS) is an interpretable reasoning framework capable of generating explicit IF-THEN fuzzy rules, making it suitable for tasks requiring transparent reasoning. However, existing ANFIS models generally construct rule antecedents and perform inference in Euclidean space, limiting their representational capacity and predictive performance. To address this issue, we propose Hyperbolic ANFIS (HyperANFIS), a hyperbolic extension of ANFIS. HyperANFIS preserves the fuzzy semantics and core architecture of conventional ANFIS while performing rule-prototype learning, rule activation, and consequent aggregation in hyperbolic space. It also retains the ability to generate interpretable IF-THEN rules. By exploiting the representational properties of hyperbolic geometry, HyperANFIS strengthens the fuzzy inference process, thereby improving predictive accuracy, inter-rule collaboration, and the credibility of its interpretable rules. Experimental results show that HyperANFIS consistently outperforms the standard ANFIS baseline and various ANFIS variants across all datasets, while also generating higher-quality fuzzy rules.

1 INTRODUCTION

Adaptive neuro-fuzzy inference systems (ANFIS) are data-adaptive machine-learning models that combine the parameter-learning capability of neural networks with the rule-based inference of fuzzy systems. ANFIS models nonlinear input-output relationships through fuzzy IF-THEN rules parameterized by membership functions (Jang, 1993). Because its rule base and rule activations can be inspected, ANFIS is particularly appealing in applications such as healthcare, industry, and finance, where transparent decision-making is important.

Despite these advantages, conventional ANFIS architectures face structural limitations as the number of input variables increases or as the data exhibit complex relational structures (Jin et al., 2024). In a classical ANFIS, the antecedent of each rule is typically formed by conjoining univariate membership functions defined on individual features in Euclidean space, with these membership functions treated independently, while the rule firing strength is obtained by aggregating their membership values. As the numbers of input variables and membership functions increase, the number of rules grows combinatorially. Meanwhile, the increasingly numerous feature representations become crowded and constrained in Euclidean space, which can affect model performance and rule-generation capacity.

This limitation is particularly pronounced when the data or decision regions exhibit hierarchical, tree-like, or highly non-uniform structures. In such cases, Euclidean geometry may provide a weak inductive bias because it does not naturally capture the expanding representational capacity required by increasingly fine-grained branches in a hierarchy (Sinha et al., 2024). Such structures are widespread across a variety of real-world applications, including healthcare and industrial systems (Koo, 2024). Therefore, it is important to develop ANFIS models in alternative geometric spaces that can enrich rule representations without sacrificing the transparency of fuzzy IF-THEN reasoning.

Hyperbolic representation learning provides a theoretically motivated geometric alternative for modeling hierarchical, tree-like, and highly non-uniform structure (Ganea et al., 2018). Owing to its

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

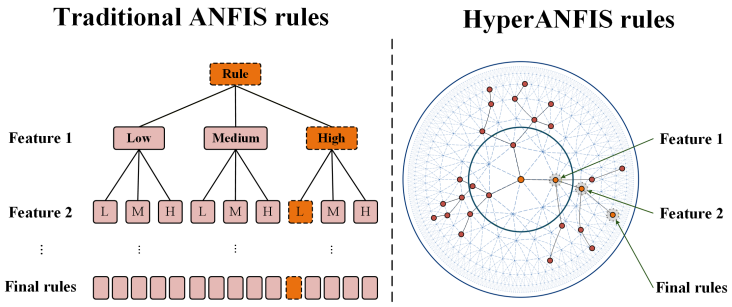

Figure 1: **Comparison of rule constraints in Euclidean and hyperbolic spaces.** The capacity of Euclidean space is limited, and the rules become severely compressed as the number of features increases. In hyperbolic space, spatial capacity grows continuously, the rules are not constrained by the available space.

constant negative curvature, the volume of a metric ball in hyperbolic space grows exponentially with its radius. This property allows a low-dimensional embedding to allocate increasing representational capacity to progressively branching hierarchies (Nickel & Kiela, 2017; Ganea et al., 2018). Prior work further suggests that hyperbolic geometry can provide a useful inductive bias when data contain explicit or latent hierarchical and non-uniform structure (Sinha et al., 2024).

To this end, we propose Hyperbolic ANFIS (HyperANFIS), a geometry-aware neuro-fuzzy model. HyperANFIS reformulates the ANFIS architecture by exploiting the properties of hyperbolic manifolds. It represents fuzzy rules using global geodesic prototypes, derives rule activation strengths from geodesic distances, and aggregates rule consequents on a hyperbolic manifold. This design preserves the scalar fuzzy semantics and functional roles of ANFIS while providing each rule with an explicit geometric interpretation. Under matched experimental settings, HyperANFIS achieves better predictive performance than its Euclidean ANFIS baseline and produces more credible rule representations.

Our contributions are as follows:

- We propose HyperANFIS, a model that learns global IF-THEN rules on a hyperbolic manifold and makes accurate predictions based on these rules.
- HyperANFIS unifies rule matching, local consequent construction, and consequent aggregation on a hyperbolic manifold, offering a new perspective on mitigating the limitations of Euclidean geometry in ANFIS.
- Experiments on multiple real-world datasets show that HyperANFIS outperforms the compared ANFIS models and produces more credible IF-THEN rules.

2 Related Work

HyperANFIS is a model that reconstructs ANFIS in hyperbolic space to derive highly interpretable rules; accordingly, research on interpretable machine learning, neuro-fuzzy systems, and hyperbolic models provides the theoretical foundation for this work.

2.1 Interpretable Machine Learning

ANFIS combine adaptive parameter learning with Takagi-Sugeno fuzzy IF-THEN inference and have long been widely used in nonlinear approximation, prediction, control, and classification tasks (Jang, 1993). Recent studies have explored different strategies to improve the scalability and interpretability of ANFIS. For example, Jin et al. simplify high-dimensional ANFIS by selecting important rules and removing similar ones (Jin et al., 2024), while UNFIS learns unstructured fuzzy rules that relax the requirement of evaluating every input variable within each rule (Salimi-Badr, 2024). Beyond rule simplification and structural relaxation, recent structure-learning approaches, such as SL-ANFIS-LSTM (Su et al., 2025), introduce adaptive fuzzy structure optimization to improve the

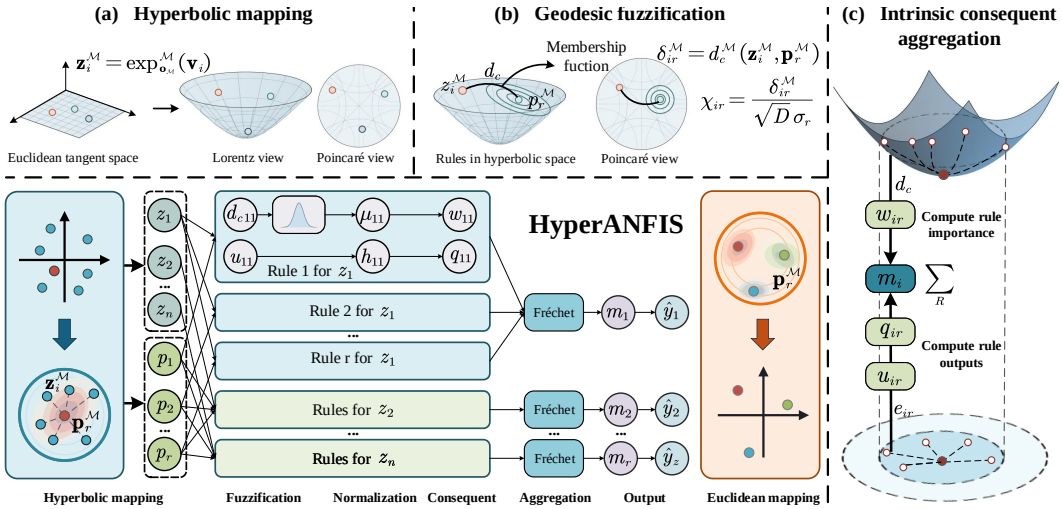

Figure 2: **Architecture of the HyperANFIS model.** The upper-left panel presents the complete model architecture. (a), (b), and (c) illustrate, respectively, the mapping of samples and rule prototypes into hyperbolic space, the generation of fuzzy rules in hyperbolic space, and the rule-based inference process.

flexibility of ANFIS-based models. More recently, KANFIS further explores a neuro-symbolic formulation based on additive function decomposition and sparse masking to reduce rule complexity and improve interpretability of fuzzy systems (Yong et al., 2026). These studies address important challenges such as rule explosion, redundancy, and antecedent complexity from different perspectives. However, existing approaches mainly focus on rule selection, structural formulation, or functional decomposition, while the geometric organization of rule representations remains largely unexplored. HyperANFIS explores the possibility of realizing ANFIS in hyperbolic space, providing a geometric perspective for improving rule representation.

2.2 ANFIS and Neuro-Fuzzy Learning

ANFIS combine adaptive parameter learning with Takagi-Sugeno fuzzy IF-THEN inference and have long been widely used in nonlinear approximation, prediction, control, and classification tasks (Jang, 1993). Recent studies on rule-base simplification, redundancy reduction, and the trade-off between predictive performance and interpretability indicate that ANFIS remains an active research area. For example, Jin et al. simplify high-dimensional ANFIS by selecting important rules and removing similar ones (Jin et al., 2024), whereas UNFIS learns unstructured fuzzy rules that do not require every input variable to be evaluated within each rule (Salimi-Badr, 2024). More recently, KANFIS introduces a neuro-symbolic formulation based on additive function decomposition and sparse masking to reduce rule complexity and improve the interpretability of fuzzy systems (Yong et al., 2026). These studies address important issues such as rule explosion, rule redundancy, and excessively complex antecedents from different perspectives. However, existing approaches mainly focus on rule selection, structural formulation, or functional decomposition, while the geometric organization of rule representations remains largely unexplored. HyperANFIS explores the possibility of realizing ANFIS in hyperbolic space, providing a geometric perspective for improving rule representation.

2.3 Theoretical Foundations and Empirical Applications of Hyperbolic Geometry

Hyperbolic geometry is theoretically well suited for representing tree-like and heterogeneous structures. In a space of constant negative curvature, the volume of a geodesic ball grows exponentially with its radius, more closely matching the branching expansion of trees than the polynomial volume growth of Euclidean space. The relationship among negative curvature, hierarchical organization,

and heterogeneous networks has been extensively studied in complex-network theory (Krioukov et al., 2010).

In representation learning, Poincaré embeddings have demonstrated that latent hierarchies can be compactly encoded in low-dimensional hyperbolic space (Nickel & Kiela, 2017); related work on hyperbolic neural networks has provided a systematic formulation of neural computation based on exponential maps, logarithmic maps, and gyrovector operations (Ganea et al., 2018), while hyperbolic graph convolutional networks have shown that such operations can preserve the hierarchical and scale-free properties of graph structures during message passing (Chami et al., 2019). Taken together, these theoretical results suggest that hyperbolic geometry may improve representation learning when the underlying task involves structured relationships. In addition, some studies have used hyperbolic objectives to encode label hierarchies, thereby reducing distortion in visual representations (Sinha et al., 2024); others have combined spherical and hyperbolic components to model cyclical homophily and hierarchical social influence in social networks, respectively (Iyer et al., 2024). In visual learning, hyperbolic visual hierarchy mapping supports classification and dense prediction tasks (Kwon et al., 2024), and hyperbolic distances have also been used in open-world object detection to capture superclass relationships (Doan et al., 2024). This line of research has further expanded to multiple modalities, with hyperbolic representations applied to compositional vision-language learning and genomic sequence modeling (Pal et al., 2025; Khan et al., 2025). Recent work has also proposed low-distortion, GPU-compatible hyperbolic tree embeddings (Van Spengler & Mettes, 2025) and investigated hyperbolic embeddings of supervised models, including decision trees (Nock et al., 2024); continuous hyperbolic latent representations have been introduced for hierarchical open-vocabulary 3D scenes (Weijler et al., 2026), and hyperbolic representation learning has been explored for reasoning over the hierarchical structure of program syntax (Zhou, 2026).

However, hyperbolization methods have not been widely studied in intrinsically interpretable models. To the best of our knowledge, this work is the first to reconstruct ANFIS in hyperbolic space, and our experiments demonstrate its effectiveness.

Overall, these studies suggest that reconstructing ANFIS in hyperbolic space has a solid research basis, promising application potential, and methodological feasibility. Building on this foundation, this work further investigates the construction and performance of HyperANFIS, with the aim of providing new insights into the structural refinement of ANFIS and the application of hyperbolic space in intrinsically interpretable models.

3 METHODOLOGY

As in Figure 2, HyperANFIS extends the neuro-fuzzy inference process to hyperbolic space, leveraging negatively curved geometry to represent fuzzy rules and perform inference. This improves the model's representational capacity and supports more discriminative, stable, and auditable fuzzy rules. This section details the architecture of HyperANFIS and its key methodological components.

3.1 OVERALL ARCHITECTURE OF HYPERANFIS

As shown in Figure 2, HyperANFIS is organized as five functional layers. Given a standardized input, the first layer maps the input and all rule centers from a shared origin tangent space to the selected hyperbolic representation, and then evaluates their intrinsic geodesic distances. The second layer converts each sample-to-rule distance into one scalar radial membership. The third layer normalizes the log-memberships across rules to obtain firing strengths. The fourth layer constructs a rule-specific consequent from the intrinsic local coordinates of the sample. The fifth layer maps these consequents to the output manifold, aggregates them through a firing-weighted Fréchet mean, and decodes the aggregate for classification or regression. All trainable parameters are optimized jointly through the resulting differentiable computation graph.

3.2 HYPERBOLIC MAPPING AND RULE PULLBACK

Let $\widehat{\mathbf{x}}_i \in \mathbf{R}^D$ denote the standardized input of sample i, and let $\mathbf{a}_r \in \mathbf{R}^D$ be the trainable tangent center of rule r. With an input scale $s > 0$ and tangent bound $\tau > 0$, HyperANFIS forms

$$\mathbf{v}_i = \mathrm{clip}_\tau(s\widehat{\mathbf{x}}_i), \qquad \overline{\mathbf{a}}_r = \mathrm{clip}_\tau(\mathbf{a}_r), \tag{1}$$

where

$$\mathrm{clip}_\tau(\mathbf{v}) = \mathbf{v}\min\left(1, \frac{\tau}{\|\mathbf{v}\|}\right). \tag{2}$$

The clipping is radial, so it preserves direction while bounding the represented geodesic radius. Define $\varrho(\mathbf{v}) = \sqrt{c}\|\mathbf{v}\|$, where $-c$ is the sectional curvature and $c > 0$.

For the Lorentz representation, the origin is $\mathbf{o}_{\mathrm{L}} = (c^{-1/2}, 0, \ldots, 0)$. The origin exponential map is

$$\exp^{\mathrm{L}}_{\mathbf{o}_{\mathrm{L}}}(\mathbf{v}) = \left(c^{-1/2}\cosh\left(\varrho(\mathbf{v})\right), \frac{\sinh\left(\varrho(\mathbf{v})\right)}{\varrho(\mathbf{v})}\mathbf{v}\right). \tag{3}$$

For the Poincaré representation, the origin is $\mathbf{o}_{\mathrm{P}} = \mathbf{0}$. Under the shared tangent convention used by HyperANFIS, its exponential map is

$$\exp^{\mathrm{P}}_{\mathbf{o}_{\mathrm{P}}}(\mathbf{v}) = \frac{\tanh\left(\varrho(\mathbf{v})/2\right)}{\varrho(\mathbf{v})}\mathbf{v}. \tag{4}$$

These maps follow the standard negative-curvature constructions used in hyperbolic representation learning (Ganea et al., 2018; Nickel & Kiela, 2018). HyperANFIS selects either $\mathcal{M} = \mathrm{L}$ or $\mathcal{M} = \mathrm{P}$ and constructs

$$\mathbf{z}_i^{\mathcal{M}} = \exp^{\mathcal{M}}_{\mathbf{o}_{\mathcal{M}}}(\mathbf{v}_i), \qquad \mathbf{p}_r^{\mathcal{M}} = \exp^{\mathcal{M}}_{\mathbf{o}_{\mathcal{M}}}(\bar{\mathbf{a}}_r). \tag{5}$$

The two representations are alternative coordinate choices and share the same trainable tangent centers.

For rule inspection, let x_{ij}^{raw} be a raw feature with training-set location ν_j and scale $\zeta_j > 0$. The preprocessing and inverse center map are

$$\begin{aligned} \widehat{x}_{ij} &= (x_{ij}^{\mathrm{raw}} - \nu_j)/\zeta_j, \\ \widetilde{a}_{rj} &= \nu_j + \zeta_j \bar{a}_{rj}/s. \end{aligned} \tag{6}$$

Thus, $\widetilde{\mathbf{a}}_r$ is a readable representative center in the original feature coordinates. It is not a coordinate-wise Euclidean rule boundary. Exact rule evaluation still replays preprocessing, clipping, hyperbolic mapping, and intrinsic distance computation, as detailed in **Appendix A**.

3.3 Geodesic Fuzzy Rule Construction

HyperANFIS defines each antecedent as one geodesic region around a global rule prototype. For Lorentz points $\mathbf{z}, \mathbf{p} \in \mathcal{H}_c^D$, define

$$\langle \mathbf{z}, \mathbf{p} \rangle_{\mathrm{L}} = -z_0 p_0 + \sum_{j=1}^{D} z_j p_j. \tag{7}$$

The corresponding intrinsic distance is

$$d_c^{\mathrm{L}}(\mathbf{z}, \mathbf{p}) = c^{-1/2}\mathrm{arcosh}\left(-c\,\langle \mathbf{z}, \mathbf{p} \rangle_{\mathrm{L}}\right). \tag{8}$$

For Poincaré points $\mathbf{z}, \mathbf{p} \in \mathcal{B}_c^D$, let $\Delta_c(\mathbf{q}) = 1 - c\|\mathbf{q}\|^2$. Their intrinsic distance is

$$d_c^{\mathrm{P}}(\mathbf{z}, \mathbf{p}) = c^{-1/2}\mathrm{arcosh}\left(1 + \frac{2c\|\mathbf{z} - \mathbf{p}\|^2}{\Delta_c(\mathbf{z})\Delta_c(\mathbf{p})}\right). \tag{9}$$

For either representation, the sample-to-rule distance is

$$\delta_{ir}^{\mathcal{M}} = d_c^{\mathcal{M}}\left(\mathbf{z}_i^{\mathcal{M}}, \mathbf{p}_r^{\mathcal{M}}\right). \tag{10}$$

Each rule has one learnable intrinsic scale. HyperANFIS parameterizes it with an unconstrained logit β_r:

$$\sigma_r = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\mathrm{sigmoid}(\beta_r). \tag{11}$$

The dimension-normalized distance and Gaussian log-membership are

$$\chi_{ir} = \frac{\delta_{ir}^{\mathcal{M}}}{\sqrt{D}\sigma_r}, \qquad \ell_{ir} = -\frac{1}{2}\chi_{ir}^2, \qquad \mu_{ir} = \exp(\ell_{ir}). \tag{12}$$

Table 1: **Performance comparison of HyperANFIS with conventional ANFIS and other ANFIS variants.** The best available result for each metric on each dataset is shown in bold. HyperANFIS achieved the best performance in all experimental results.

Model	Spambase			Car			Zoo			WDBC			NSL-KDD		
	Acc.	F1	Recall	Acc.	F1	Recall	Acc.	F1	Recall	Acc.	F1	Recall	Acc.	F1	Recall
FSRE-AdaTSK	0.7742	0.7544	0.7477	0.5896	0.2619	0.2811	0.3810	0.2379	0.3095	0.7544	0.7304	0.7262	0.7192	0.4235	0.4384
FCM-ANFIS	0.9229	0.9198	0.9229	0.8815	0.8062	0.9076	0.7142	0.6920	0.7797	0.9035	0.8892	0.9712	0.7932	0.5878	0.5910
IT2-ANFIS	0.9121	0.9088	0.9130	0.9104	0.8469	0.9086	0.8095	0.5946	0.6607	0.9123	0.9037	0.8958	0.7622	0.5856	0.5842
PSO-ANFIS	0.9044	0.8996	0.8985	0.8150	0.6517	0.9364	0.9047	0.7809	0.8095	0.8947	0.8889	0.8968	0.7620	0.5602	0.5485
ANFIS	0.9120	0.9080	0.9075	0.8998	0.8277	0.8731	0.8413	0.6875	0.7500	0.9152	0.9067	0.8958	0.7798	0.5952	0.5935
HyperANFIS (Ours)	**0.9251**	**0.9215**	**0.9251**	**0.9181**	**0.8827**	**0.9449**	**0.9524**	**0.8186**	**0.8571**	**0.9854**	**0.9843**	**0.9835**	**0.8038**	**0.6709**	**0.6281**

Consequently, $\sqrt{D}\sigma_r$ is the geodesic radius at which $\mu_{ir} = \exp(-1/2)$. The implementation also supports a generalized Bell kernel, whose exact form is given in **Appendix B**.

The third layer normalizes the log-memberships directly:

$$\overline{w}_{ir} = \frac{\exp(\ell_{ir})}{\sum_{q=1}^{R} \exp(\ell_{iq})}. \tag{13}$$

This construction yields one scalar activation for each sample-rule pair. It therefore represents a global geodesic fuzzy rule rather than a product of independent coordinate-wise memberships.

3.4 Intrinsic Consequent Aggregation

Let $H = \max(2, O)$ be the output-manifold dimension, where O is the task output dimension. HyperANFIS first expresses sample i relative to rule prototype r through the intrinsic local coordinate

$$\mathbf{e}_{ir} = \left[\mathcal{T}_{\mathbf{p}_r^{\mathcal{M}} \to \mathbf{o}_{\mathcal{M}}} \left(\log_{\mathbf{p}_r^{\mathcal{M}}}^{\mathcal{M}} (\mathbf{z}_i^{\mathcal{M}}) \right) \right]_{\mathrm{sp}}, \tag{14}$$

where $\mathcal{T}$ denotes parallel transport and $[\cdot]_{\mathrm{sp}}$ retains the D spatial tangent coordinates. A first-order rule consequent is

$$\begin{aligned} \mathbf{u}_{ir} &= \mathrm{clip}_{\tau} \left(\mathbf{b}_r + \mathbf{W}_r \mathbf{e}_{ir} \right), \\ \mathbf{q}_{ir}^{\mathcal{M}} &= \exp_{\mathbf{o}_{\mathcal{M}}}^{\mathcal{M}} (\mathbf{u}_{ir}). \end{aligned} \tag{15}$$

where $\mathbf{b}_r \in \mathbf{R}^H$ and $\mathbf{W}_r \in \mathbf{R}^{H \times D}$. The zero-order variant omits $\mathbf{W}_r \mathbf{e}_{ir}$.

The fifth layer aggregates the manifold-valued consequents through their weighted Fréchet mean (Lou et al., 2020):

$$\mathbf{m}_i^{\mathcal{M}} = \operatorname*{argmin}_{\mathbf{m} \in \mathcal{M}} \sum_{r=1}^{R} \overline{w}_{ir} d_c^{\mathcal{M}} (\mathbf{m}, \mathbf{q}_{ir}^{\mathcal{M}})^2, \tag{16}$$

HyperANFIS computes this mean by a fixed number of differentiable Karcher refinements. For classification, trainable class tangents $\mathbf{g}_k$ define class prototypes $\mathbf{c}_k^{\mathcal{M}} = \exp_{\mathbf{o}_{\mathcal{M}}}^{\mathcal{M}} (\mathrm{clip}_{\tau}(\mathbf{g}_k))$, and the class logits are

$$s_{ik} = -d_c^{\mathcal{M}} (\mathbf{m}_i^{\mathcal{M}}, \mathbf{c}_k^{\mathcal{M}})^2. \tag{17}$$

For regression, the prediction is obtained in the origin tangent chart:

$$\widehat{\mathbf{y}}_i = \left[\log_{\mathbf{o}_{\mathcal{M}}}^{\mathcal{M}} (\mathbf{m}_i^{\mathcal{M}}) \right]_{1:O}. \tag{18}$$

3.5 Forward Computation

For a standardized input $\mathbf{x}_i$, the complete HyperANFIS forward computation is summarized as

$$\begin{aligned} \left\{ (\overline{w}_{ir}, \mathbf{q}_{ir}^{\mathcal{M}}) \right\}_{r=1}^{R} &= \mathcal{R}_{\Theta, \mathcal{M}} (\mathbf{x}_i), \\ \mathbf{m}_i^{\mathcal{M}} &= \operatorname*{argmin}_{\mathbf{m} \in \mathcal{M}} \sum_{r=1}^{R} \overline{w}_{ir} d_c^{\mathcal{M}} (\mathbf{m}, \mathbf{q}_{ir}^{\mathcal{M}})^2, \\ \widehat{\mathbf{y}}_i &= \mathcal{D} (\mathbf{m}_i^{\mathcal{M}}). \end{aligned} \tag{19}$$

Table 2: **Selected HyperANFIS rules learned on the WDBC dataset.** Each row represents a single IF-THEN rule together with the corresponding support from the literature. An individual rule is not the prediction for a single sample; rather, its firing strength is influenced by its dominant coverage.

Rule	Projected, interpretable IF-THEN rule	Morphological evidence
R4	**IF** *mean compactness* **is HIGH** ($\widetilde{c} = +3.69$), *mean concavity* **is HIGH** ($\widetilde{c} = +3.65$), *mean concave points* **is HIGH** ($\widetilde{c} = +3.77$), and *worst concave points* **is HIGH** ($\widetilde{c} = +3.84$). **THEN** the rule-level conclusion is **malignant (M)**. *Dominant coverage: 26.3%.*	High compactness, concavity, and concave-point counts describe pronounced nuclear-boundary indentation and irregularity. Direct breast-aspirate morphometry found compactness and concave-point counts to be significantly higher in ductal carcinoma than in benign lesions (Narasimha et al., 2013).
R10	**IF** *mean radius* **is LOW** ($\widetilde{c} = -2.24$), *mean concavity* **is LOW** ($\widetilde{c} = -1.98$), *mean concave points* **is LOW** ($\widetilde{c} = -2.18$), and *worst radius* **is LOW** ($\widetilde{c} = -2.37$). **THEN** the rule-level conclusion is **benign (B)**. *Dominant coverage: 21.1%.*	Small nuclei with low concavity and few concave points form a compact, comparatively smooth nuclear phenotype. Direct breast-aspirate morphometry reported lower nuclear size, compactness, and concave-point counts in benign lesions than in ductal carcinoma (Narasimha et al., 2013).
R2	**IF** *mean radius* **is LOW** ($\widetilde{c} = -1.11$), *mean area* **is LOW** ($\widetilde{c} = -1.21$), *worst radius* **is LOW** ($\widetilde{c} = -1.32$), and *worst area* **is LOW** ($\widetilde{c} = -0.70$). **THEN** the rule-level conclusion is **benign (B)**. *Dominant coverage: 14.9%.*	Low mean and worst-case radius and area indicate that both typical and extreme nuclear sizes remain small. This size-dominant pattern agrees with cytological evidence showing lower nuclear area, diameter, and perimeter in benign breast aspirates (Pandian et al., 2021).
R6	**IF** *mean concavity* **is LOW** ($\widetilde{c} = -2.06$), *radius SE* **is LOW** ($\widetilde{c} = -2.23$), *perimeter SE* **is LOW** ($\widetilde{c} = -2.12$), and *worst radius* **is LOW** ($\widetilde{c} = -2.37$). **THEN** the rule-level conclusion is **benign (B)**. *Dominant coverage: 12.3%.*	Low concavity and worst radius support a compact, regular nuclear phenotype; low radius and perimeter SE indicate limited within-sample size dispersion. This direction agrees with the greater nuclear size and size dispersion reported in malignant aspirates (Pandian et al., 2021).

Here, $\mathcal{R}_{\Theta,\mathcal{M}}$ denotes hyperbolic rule inference and produces the normalized firing strength $\overline{w}_{ir}$ and intrinsic consequent $\mathbf{q}_{ir}^{\mathcal{M}}$ for each rule. These consequents are then aggregated through their activation-weighted Fréchet mean, and $\mathcal{D}$ decodes the resulting manifold point into the task prediction. Detailed definitions and optimization procedures are provided in the **Appendix C**.

4 Experiments

To systematically evaluate whether hyperbolic geometric modeling improves the predictive performance and rule interpretability of ANFIS, we compare HyperANFIS with conventional ANFIS and other representative ANFIS variants across multiple datasets. The selected datasets cover several application domains, including healthcare and finance, and vary in complexity, thereby providing a broad experimental basis for model evaluation. In addition, we use the Wisconsin Diagnostic Breast Cancer (WDBC) dataset as a case study and analyze the generated rules from multiple perspectives, providing further evidence of the improved rule interpretability achieved by HyperANFIS.

4.1 Comparison Experiments

We compared HyperANFIS with conventional ANFIS, which served as the primary Euclidean baseline, and four representative neuro-fuzzy models. Conventional ANFIS adopts a dimension-wise Takagi-Sugeno inference structure and learns its antecedent and consequent parameters from the training data (Zhang & Chen, 2024). FSRE-AdaTSK integrates feature selection, rule extraction, and parameter fine-tuning within an adaptive TSK fuzzy system (Xue et al., 2022). FCM-ANFIS uses fuzzy c-means clustering to initialize fuzzy partitions and rule prototypes (Knaiber & Alawieh, 2023). In our implementation, these prototypes were subsequently refined through gradient-based optimization. IT2-ANFIS represents uncertainty in rule membership by introducing lower and upper interval type-2 membership bounds into its coordinate-wise rule antecedents (Zand et al., 2024). In our implementation, a midpoint approximation was obtained by averaging the lower and upper

bounds of the corresponding rule firing strengths. Finally, PSO-ANFIS uses particle swarm optimization to optimize the trainable parameters of ANFIS (Vesović et al., 2024). In our implementation, PSO provided a supervised warm start for all trainable parameters, followed by a common Adam-based refinement stage.

The five datasets cover email spam detection (Spambase), vehicle acceptability evaluation (Car), animal category classification (Zoo), breast cancer diagnosis (WDBC), and network intrusion detection (NSL-KDD). They encompass both binary and multiclass classification tasks and differ in sample size, feature dimensionality, feature characteristics, and class structure, providing a diverse test bed for evaluating model performance across application domains and data distributions. For each random seed, all methods used consistent data preprocessing, data partitions, and evaluation metrics, and the reported results were averaged over five predefined random seeds.

As shown in Table 1, HyperANFIS outperforms conventional ANFIS and the representative neuro-fuzzy baselines on all five datasets. Compared with conventional ANFIS, HyperANFIS improves the average accuracy, Macro-F1 score, and Recall by 0.0473, 0.0706, and 0.0638, respectively. As discussed in the following sections, HyperANFIS also yields more reliable predictions than the comparison models in terms of the output rule attributes. These results demonstrate that geometry-aware rule inference can substantially enhance the predictive capability of ANFIS and indicate that hyperbolic computation is a promising direction for further improving ANFIS and its variants.

These improvements are particularly pronounced on Zoo and WDBC. On the Zoo dataset, HyperANFIS improves accuracy, Macro-F1, and Recall by 11.11, 13.11, and 0.1071 percentage points, respectively; on WDBC, the corresponding improvements reach 7.02, 7.76, and 0.0877 percentage points. Zoo exhibits a natural hierarchical taxonomic structure, whereas WDBC contains complex and nonuniform diagnostic relationships among its features. These consistent improvements on the two datasets provide empirical support for the effectiveness of negative-curvature geometry in organizing nonuniform and hierarchical rule relationships, thereby leading to performance gains.

4.2 Interpretability Analysis

To evaluate whether hyperbolic geometry improves the rule interpretability of ANFIS, we compared HyperANFIS with classical ANFIS on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and analyzed representative IF-THEN rules. WDBC contains 569 instances and 30 real-valued features extracted from digitized images of fine-needle aspirates of breast masses. These features describe nuclear size, texture, and contour morphology, providing explicit cytomorphological semantics for evaluating rule utilization, distinctiveness, and morphological consistency.

As shown in Table 2, HyperANFIS preserves the ability of classical ANFIS to express its inference process through interpretable IF-THEN rules. For readability, the table reports the four rules with the highest dominant coverage on the validation set and retains only three to four representative features for each rule, while the omitted features remain part of the complete antecedent. The learned rules associate malignancy with increased nuclear size and contour irregularity, whereas benign patterns are generally characterized by smaller nuclear size and lower concavity. These patterns are consistent with established findings in breast nuclear morphometry.

4.3 Comparison with Classical ANFIS

We compare the rule characteristics learned by conventional ANFIS and HyperANFIS on the WDBC dataset, which has a complex feature space and strong intrinsic logical structure. As shown in Figure 3, the interpretability of conventional ANFIS deteriorates substantially on this dataset, where the feature complexity and underlying relational structure make rule learning challenging. In contrast, HyperANFIS preserves a high level of interpretability, revealing a clear distinction between the two models. This experiment was conducted with a fixed seed under fair comparison conditions.

In Figure 3(a), classical ANFIS assigns 42.98% of the validation samples to R2 as their dominant rule, whereas the dominant coverage of the second most frequently selected rule drops to 10.53%. In contrast, HyperANFIS distributes dominant coverage across five principal rules: R4 (26.32%), R10 (21.05%), R2 (14.91%), R6 (12.28%), and R9 (12.28%). The remaining rules also participate in inference through nonzero mean firing strengths. Thus, HyperANFIS transforms the single-rule-dominated utilization pattern of classical ANFIS into a broader and more clearly organized multi-

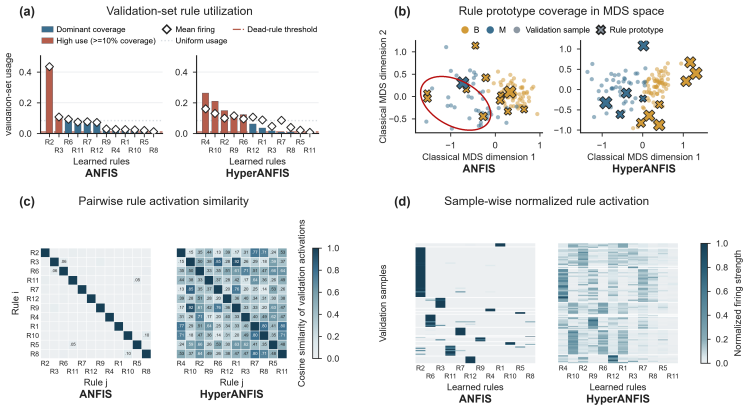

Figure 3: **Comparison of the interpretability of ANFIS and HyperANFIS on the WDBC dataset.** The ANFIS results are shown on the left, whereas the HyperANFIS results are shown on the right. (a) shows the frequency with which each rule is used for prediction, (b) visualizes the relationship between the rules and the dataset, with red circles marking the centers of rules associated with severe errors, (c) shows the correlations among the rules, and (d) illustrates the contribution of each rule to each individual prediction.

rule participation pattern. Figure 3(b) further demonstrates this improvement at the rule-prototype level. Classical ANFIS learns 11 benign prototypes but only one malignant prototype, resulting in a substantial imbalance in class representation. In contrast, HyperANFIS learns seven benign and five malignant prototypes, whose distribution is more coherent. As indicated by the red circles in the figure, some rule prototypes are also clearly mislearned. Figure 3(c) shows that the mean off-diagonal cosine similarity between rule activation vectors increases from 0.006 to 0.432. The rules learned by classical ANFIS are activated in an almost completely mutually exclusive manner, which appears difficult to reconcile with the principles of fuzzy inference. HyperANFIS, in contrast, establishes correlated yet differentiated relationships among its rules, thereby better conforming to the gradual evidence-combination process underlying fuzzy reasoning. Figure 3(d) provides direct sample-level evidence: the mean entropy-derived effective number of active rules increases from 1.14 for classical ANFIS to 6.67 for HyperANFIS. Classical ANFIS consequently exhibits an almost winner-take-all inference pattern dominated by a single rule, whereas HyperANFIS uses one or more principal rules to guide the inference, with the remaining rules providing supplementary evidence.

These results indicate that hyperbolic geometry enables the learned IF-THEN rules to provide more complete class representations, stronger inter-rule cooperation, and a more trustworthy inference process. Together with the accuracy results reported in Table 1, the findings show that HyperANFIS achieves higher accuracy and more trustworthy predictions while producing higher-quality rules.

5 Conclusion

This work investigated how the geometric properties of hyperbolic space can improve ANFIS, an intrinsically interpretable model. By reformulating rule prototypes, rule activation, and consequent aggregation in hyperbolic space while preserving the original IF-THEN semantics, HyperANFIS consistently outperformed classical ANFIS and representative ANFIS variants in accuracy and Macro-F1 across the five evaluated datasets. The WDBC analysis further demonstrated that HyperANFIS learned a more class-diverse, cooperative, and auditable rule system than classical ANFIS. The significance of this work therefore extends beyond the geometric reformulation and performance improvement of classical ANFIS. It provides a new geometric perspective for improving other ANFIS-family models and, more broadly, intrinsically interpretable models.

AI USE STATEMENT

During the preparation of this work, we used AI tools to assist with writing and polishing the manuscript, retrieving and identifying relevant literature, and developing proofs of mathematical claims presented in the paper. The authors take full responsibility for the accuracy and integrity of all content, including the cited literature, mathematical claims, and proofs.

REFERENCES

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

Thang Doan, Xin Li, Sima Behpour, Wenbin He, Liang Gou, and Liu Ren. Hyp-ow: Exploiting hierarchical structure learning with hyperbolic distance enhances open world object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1555–1563, 2024.

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

Roshni G Iyer, Yewen Wang, Wei Wang, and Yizhou Sun. Non-euclidean mixture model for social network embedding. *Advances in Neural Information Processing Systems*, 37:111464–111488, 2024.

J.-S.R. Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993. doi: 10.1109/21.256541.

Yali Jin, Weihua Cao, Min Wu, Yan Yuan, and Yang Shi. Simplification of anfis based on importance-confidence-similarity measures. *Fuzzy Sets Syst.*, 481(C), April 2024. ISSN 0165-0114. doi: 10.1016/j.fss.2024.108887. URL https://doi.org/10.1016/j.fss.2024.108887.

Raiyan Khan, Philippe Chlenski, and Itsik Pe'er. Hyperbolic genome embeddings. In *International Conference on Learning Representations*, volume 2025, pp. 73425–73454, 2025.

Mohammed Knaiber and Leen Alawieh. Bayesian inference using an adaptive neuro-fuzzy inference system. *Fuzzy Sets and Systems*, 459:43–66, 2023.

Heejoon Koo. Next visit diagnosis prediction via medical code-centric multimodal contrastive EHR modelling with hierarchical regularisation. In Yvette Graham and Matthew Purver (eds.), *Findings of the Association for Computational Linguistics: EACL 2024*, pp. 41–55, St. Julian's, Malta, March 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-eacl.3. URL https://aclanthology.org/2024.findings-eacl.3/.

Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(3):036106, 2010.

Hyeongjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, and Kwanghoon Sohn. Improving visual recognition with hyperbolical visual hierarchy mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17364–17374, 2024.

Aaron Lou, Isay Katsman, Qingxuan Jiang, Serge Belongie, Ser-Nam Lim, and Christopher De Sa. Differentiating through the fréchet mean. In *International conference on machine learning*, pp. 6393–6403. PMLR, 2020.

Aparna Narasimha, B Vasavi, and Harendra ML Kumar. Significance of nuclear morphometry in benign and malignant breast aspirates. *International Journal of Applied and Basic Medical Research*, 3(1):22–26, 2013.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/59dfa2df42d9e3d41f5b02bfc32229dd-Paper.pdf.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pp. 3779–3788. PMLR, 2018.

Richard Nock, Ehsan Amid, Frank Nielsen, Alexander Soen, and Manfred K Warmuth. Hyperbolic embeddings of supervised models. *Advances in Neural Information Processing Systems*, 37: 140865–140902, 2024.

Avik Pal, Max van Spengler, D Guido Maria, Amely di Melendugno, Alessandro Flaborea, Fabio Galasso, and Pascal Mettes. Compositional entailment learning for hyperbolic vision-language models (2024), 2025.

Dayal Johan Niranjan Pandian, Anita Ramdas, and M Moses Ambroise. Image analysis-assisted nuclear morphometric study of benign and malignant breast aspirates. *Journal of Microscopy and Ultrastructure*, 9(3):114–118, 2021.

Armin Salimi-Badr. Unfis: A novel neuro-fuzzy inference system with unstructured fuzzy rules. *Neurocomputing*, 579:127437, 2024.

Aditya Sinha, Siqi Zeng, Makoto Yamada, and Han Zhao. Learning structured representations with hyperbolic embeddings. *Advances in Neural Information Processing Systems*, 37:91220–91259, 2024.

Zhao Su, Jun Shen, Qingguo Zhou, and Binbin Yong. Sl-anfis-lstm: A structure learnable fuzzy neural network for ultra-short-term pv power forecasting. *IEEE Transactions on Fuzzy Systems*, 2025.

Max Van Spengler and Pascal Mettes. Low-distortion and gpu-compatible tree embeddings in hyperbolic space. *arXiv preprint arXiv:2502.17130*, 2025.

Mitra Vesović, Radiša Jovanović, and Vladimir Zarić. Hybrid ga-anfis and pso-anfis techniques for nonlinear dc motor system modeling. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, pp. 09544062251350800, 2024.

Lisa Weijler, Sebastian Koch, Fabio Poiesi, Timo Ropinski, and Pedro Hermosilla. Openhype: Hyperbolic embeddings for hierarchical open-vocabulary radiance fields. *Advances in Neural Information Processing Systems*, 38:119039–119068, 2026.

Guangdong Xue, Qin Chang, Jian Wang, Kai Zhang, and Nikhil Ranjan Pal. An adaptive neuro-fuzzy system with integrated feature selection and rule extraction for high-dimensional classification problems. *IEEE Transactions on Fuzzy Systems*, 31(7):2167–2181, 2022.

Binbin Yong, Haoran Pei, Jun Shen, Haoran Li, Qingguo Zhou, and Zhao Su. Kanfis a neuro-symbolic framework for interpretable and uncertainty-aware learning. *arXiv preprint arXiv:2602.03034*, 2026.

Javad Palizvan Zand, Javad Katebi, and Saman Yaghmaei-Sabegh. A hybrid clustering-based type-2 adaptive neuro-fuzzy forecasting model for smart control systems. *Expert Systems with Applications*, 239:122445, 2024.

Dongsong Zhang and Tianhua Chen. Scikit-anfis: A scikit-learn compatible python implementation for adaptive neuro-fuzzy inference system. *Int. J. Fuzzy Syst.*, 26(6):2039–2057, 2024.

Weilin Zhou. The Natural Geometry of Code: Hyperbolic Representation Learning for Program Reasoning. In *International Conference on Learning Representations*, 2026. URL `https://mlanthology.org/iclr/2026/zhou2026iclr-natural/`.

This supplement provides additional geometric definitions, model variants, optimization details, theoretical analysis, and implementation settings for HyperANFIS.

A HYPERBOLIC GEOMETRY

We define the Lorentz and Poincaré operations under a shared tangent-space parameterization. Throughout, $c > 0$ denotes the curvature magnitude, and the sectional curvature is $-c$. Input scaling, projection, and finite-precision safeguards are specified in Appendix E.

A.1 TANGENT-SPACE PARAMETERIZATION

Let $\widehat{\mathbf{x}}_i \in \mathbb{R}^D$ denote a standardized input and let $\mathbf{a}_r \in \mathbb{R}^D$ denote the trainable tangent-space center of rule r. HyperANFIS parameterizes samples and rule centers as

$$\mathbf{v}_i = \mathrm{clip}_\tau(s\widehat{\mathbf{x}}_i), \qquad \overline{\mathbf{a}}_r = \mathrm{clip}_\tau(\mathbf{a}_r), \tag{20}$$

where $s > 0$ is the input scale and $\tau > 0$ is the maximum tangent norm. The clipping operator is

$$\mathrm{clip}_\tau(\mathbf{v}) = \begin{cases} \mathbf{0}, & \|\mathbf{v}\| = 0, \\ \mathbf{v}\min\left(1, \dfrac{\tau}{\|\mathbf{v}\|}\right), & \|\mathbf{v}\| > 0. \end{cases} \tag{21}$$

Radial clipping preserves the direction of each nonzero tangent vector and bounds the corresponding geodesic radius. The definition at zero gives the continuous extension. The data-dependent scale s is specified in Appendix E.1.

A.2 LORENTZ MODEL

The D-dimensional Lorentz model is the upper sheet of the hyperboloid

$$\mathcal{H}_c^D = \left\{\mathbf{z} \in \mathbb{R}^{D+1} : \langle \mathbf{z}, \mathbf{z}\rangle_{\mathrm{L}} = -c^{-1},\ z_0 > 0\right\}, \tag{22}$$

with Lorentz inner product

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\mathrm{L}} = -x_0 y_0 + \sum_{j=1}^{D} x_j y_j. \tag{23}$$

Its tangent space at $\mathbf{x} \in \mathcal{H}_c^D$ is

$$T_{\mathbf{x}}\mathcal{H}_c^D = \left\{\mathbf{u} \in \mathbb{R}^{D+1} : \langle \mathbf{x}, \mathbf{u}\rangle_{\mathrm{L}} = 0\right\}. \tag{24}$$

For $\mathbf{u} \in T_{\mathbf{x}}\mathcal{H}_c^D$, define $\|\mathbf{u}\|_{\mathrm{L}} = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle_{\mathrm{L}}}$. The exponential map is

$$\begin{aligned} \exp_{\mathbf{x}}^{\mathrm{L}}(\mathbf{u}) = \cosh\left(\sqrt{c}\|\mathbf{u}\|_{\mathrm{L}}\right)\mathbf{x} \\ + \frac{\sinh\left(\sqrt{c}\|\mathbf{u}\|_{\mathrm{L}}\right)}{\sqrt{c}\|\mathbf{u}\|_{\mathrm{L}}}\mathbf{u}. \end{aligned} \tag{25}$$

At the origin $\mathbf{o}_{\mathrm{L}} = (c^{-1/2}, 0, \ldots, 0)$, a shared tangent $\mathbf{v} \in \mathbb{R}^D$ is identified with $(0, \mathbf{v}) \in T_{\mathbf{o}_{\mathrm{L}}}\mathcal{H}_c^D$. With $\varrho(\mathbf{v}) = \sqrt{c}\|\mathbf{v}\|$, Equation (25) becomes

$$\begin{aligned} \exp_{\mathbf{o}_{\mathrm{L}}}^{\mathrm{L}}(\mathbf{v}) = \Bigg(c^{-1/2}\cosh(\varrho(\mathbf{v})), \\ \frac{\sinh(\varrho(\mathbf{v}))}{\varrho(\mathbf{v})}\mathbf{v}\Bigg). \end{aligned} \tag{26}$$

The spatial multiplier is defined by its continuous limit, 1, at the origin. Direct substitution shows that the image satisfies Equation (22).

For $\mathbf{x}, \mathbf{y} \in \mathcal{H}_c^D$, let

$$\alpha(\mathbf{x}, \mathbf{y}) = -c\langle \mathbf{x}, \mathbf{y}\rangle_{\mathrm{L}} \geq 1. \tag{27}$$

The geodesic distance and logarithmic map are

$$d_c^{\mathrm{L}}(\mathbf{x},\mathbf{y}) = c^{-1/2}\mathrm{arcosh}(\alpha), \tag{28}$$

$$\log_{\mathbf{x}}^{\mathrm{L}}(\mathbf{y}) = \frac{\mathrm{arcosh}(\alpha)}{\sqrt{\alpha^2-1}}(\mathbf{y}-\alpha\mathbf{x}). \tag{29}$$

The direction in Equation (29) is tangent at $\mathbf{x}$ and has squared Lorentz norm $(\alpha^2-1)/c$; consequently, $\|\log_{\mathbf{x}}^{\mathrm{L}}(\mathbf{y})\|_{\mathrm{L}} = d_c^{\mathrm{L}}(\mathbf{x},\mathbf{y})$, which verifies the scale factor in Equation (29).

For $\mathbf{u} \in T_{\mathbf{x}}\mathcal{H}_c^D$, parallel transport along the unique geodesic from $\mathbf{x}$ to $\mathbf{y}$ is

$$\mathcal{T}_{\mathbf{x}\to\mathbf{y}}^{\mathrm{L}}(\mathbf{u}) = \mathbf{u} + \frac{\langle \mathbf{y},\mathbf{u}\rangle_{\mathrm{L}}}{c^{-1}-\langle \mathbf{x},\mathbf{y}\rangle_{\mathrm{L}}}(\mathbf{x}+\mathbf{y}). \tag{30}$$

The denominator is $(1+\alpha)/c > 0$. Parallel transport preserves the Lorentz norm and maps the vector to $T_{\mathbf{y}}\mathcal{H}_c^D$. HyperANFIS uses Equation (30) with $\mathbf{y} = \mathbf{o}_{\mathrm{L}}$ to express all rule-local coordinates in the shared origin chart.

A.3 Poincaré Ball and Lorentz Isometry

The Poincaré ball of curvature $-c$ is

$$\mathcal{B}_c^D = \left\{\mathbf{q} \in \mathbb{R}^D : c\|\mathbf{q}\|^2 < 1\right\}, \tag{31}$$

with conformal metric

$$g_{\mathbf{q}}^{\mathrm{P}} = \lambda_{\mathbf{q}}^2 g^{\mathrm{E}}, \qquad \lambda_{\mathbf{q}} = \frac{2}{1-c\|\mathbf{q}\|^2}. \tag{32}$$

For $\mathbf{z},\mathbf{p} \in \mathcal{B}_c^D$, define $\Delta_c(\mathbf{q}) = 1-c\|\mathbf{q}\|^2$. Their distance is

$$d_c^{\mathrm{P}}(\mathbf{z},\mathbf{p}) = c^{-1/2}\mathrm{arcosh}\left(1 + \frac{2c\|\mathbf{z}-\mathbf{p}\|^2}{\Delta_c(\mathbf{z})\Delta_c(\mathbf{p})}\right). \tag{33}$$

Both hyperbolic models use the same tangent vector at the origin. Under this convention, the Poincaré exponential map is

$$\exp_{\mathbf{o}_{\mathrm{P}}}^{\mathrm{P}}(\mathbf{v}) = \frac{\tanh(\varrho(\mathbf{v})/2)}{\varrho(\mathbf{v})}\mathbf{v}, \qquad \mathbf{o}_{\mathrm{P}} = \mathbf{0}. \tag{34}$$

Its differential at the origin maps $\mathbf{v}$ to $\mathbf{v}/2$. Because $\lambda_{\mathbf{0}} = 2$, the Riemannian norm of this derivative equals the Euclidean norm of the shared tangent vector $\mathbf{v}$.

The isometry from the Poincaré ball to the Lorentz model is

$$\Phi_c(\mathbf{q}) = \left(\frac{1+c\|\mathbf{q}\|^2}{\sqrt{c}(1-c\|\mathbf{q}\|^2)}, \frac{2\mathbf{q}}{1-c\|\mathbf{q}\|^2}\right), \tag{35}$$

with inverse

$$\Phi_c^{-1}(\mathbf{z}) = \frac{\mathbf{z}_{\mathrm{sp}}}{\sqrt{c}z_0+1}. \tag{36}$$

Using the standard hyperbolic half-angle identities gives

$$\Phi_c\left(\exp_{\mathbf{o}_{\mathrm{P}}}^{\mathrm{P}}(\mathbf{v})\right) = \exp_{\mathbf{o}_{\mathrm{L}}}^{\mathrm{L}}(\mathbf{v}). \tag{37}$$

Direct substitution into the Lorentz inner product yields

$$-c\left\langle \Phi_c(\mathbf{z}),\Phi_c(\mathbf{p})\right\rangle_{\mathrm{L}} = 1 + \frac{2c\|\mathbf{z}-\mathbf{p}\|^2}{\Delta_c(\mathbf{z})\Delta_c(\mathbf{p})}. \tag{38}$$

Equations (28)– (38) imply $d_c^{\mathrm{P}}(\mathbf{z},\mathbf{p}) = d_c^{\mathrm{L}}(\Phi_c(\mathbf{z}),\Phi_c(\mathbf{p}))$. Accordingly, Poincaré distances and tangent-space operations are evaluated after mapping points through Φ_c. Manifold-valued aggregation results are mapped back through Φ_c^{-1}. These computations represent the same intrinsic operations in two coordinate systems.

B GEOMETRY AND MEMBERSHIP VARIANTS

HyperANFIS supports one geometry and one membership function per training run. The variants share the rule architecture and parameterization; multiple geometries or membership functions are not combined within a forward pass.

B.1 GEOMETRY VARIANTS

The Lorentz variant applies Equations (22)– (30) directly and is the default. The Poincaré variant applies Equations (31)– (38); its bounded coordinates represent the same intrinsic distances as the Lorentz model.

The Euclidean variant preserves the rule architecture but replaces the manifold operations by

$$\begin{aligned} \exp^{\mathrm{E}}_{\mathbf{o}_{\mathrm{E}}}(\mathbf{v}) &= \mathbf{v}, \\ d^{\mathrm{E}}(\mathbf{x},\mathbf{y}) &= \|\mathbf{x}-\mathbf{y}\|, \\ \mathbf{e}^{\mathrm{E}}_{ir} &= \mathbf{z}^{\mathrm{E}}_{i} - \mathbf{p}^{\mathrm{E}}_{r}, \\ \mathbf{m}^{\mathrm{E}}_{i} &= \sum_{r=1}^{R} \overline{w}_{ir}\mathbf{q}^{\mathrm{E}}_{ir}. \end{aligned} \tag{39}$$

The Euclidean and hyperbolic variants therefore use the same number of rules, consequent parameterization, and firing normalization. The Euclidean variant is a geometry-matched counterpart of HyperANFIS, not a conventional coordinate-wise ANFIS.

B.2 MEMBERSHIP FUNCTIONS

For any geometry, define the dimension-normalized intrinsic distance as

$$\chi_{ir} = \frac{\delta^{\mathcal{M}}_{ir}}{\sqrt{D}\sigma_r}. \tag{40}$$

Each rule scale is bounded through the logit parameterization

$$\sigma_r = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\mathrm{sigmoid}(\beta_r). \tag{41}$$

The Gaussian membership function is

$$\ell^{\mathrm{G}}_{ir} = -\frac{1}{2}|\chi_{ir}|^2, \qquad \mu^{\mathrm{G}}_{ir} = \exp(\ell^{\mathrm{G}}_{ir}). \tag{42}$$

Consequently, the intrinsic radius $\sqrt{D}\sigma_r$ corresponds to $\mu^{\mathrm{G}}_{ir} = \exp(-1/2)$.

The generalized Bell membership function with fixed $b > 0$ is

$$\ell^{\mathrm{B}}_{ir} = -\log\left(1 + |\chi_{ir}|^{2b}\right), \qquad \mu^{\mathrm{B}}_{ir} = \frac{1}{1+|\chi_{ir}|^{2b}}. \tag{43}$$

Both functions equal 1 at the rule center. The Gaussian tail decays exponentially in χ^2_{ir}, whereas the Bell tail decays polynomially as $|\chi_{ir}|^{-2b}$. The Bell shape parameter b controls the transition between the central region and the tail; the default is $b = 2$.

Normalized firing strengths are computed in the log domain:

$$\begin{aligned} \overline{w}_{ir} &= \frac{\exp(\ell_{ir})}{\sum_{q=1}^{R}\exp(\ell_{iq})} \\ &= \exp\left(\ell_{ir} - \mathrm{logsumexp}_q(\ell_{iq})\right). \end{aligned} \tag{44}$$

This form does not require explicit exponentiation before normalization and therefore avoids underflow of intermediate membership values. Numerical safeguards are specified in Appendix E.2.

C Inference and Optimization

This section specifies rule-local consequent construction, intrinsic aggregation, the training objective, and gradient computation.

C.1 Rule-Local Consequents

For $\mathcal{M} \in \{\mathrm{L}, \mathrm{P}\}$, the local coordinate of sample i relative to rule r is

$$\begin{aligned} \boldsymbol{\xi}_{ir} &= \log^{\mathcal{M}}_{\mathbf{p}_r^{\mathcal{M}}}(\mathbf{z}_i^{\mathcal{M}}), \\ \boldsymbol{\eta}_{ir} &= \mathcal{T}^{\mathcal{M}}_{\mathbf{p}_r^{\mathcal{M}} \to \mathbf{o}_{\mathcal{M}}}(\boldsymbol{\xi}_{ir}), \\ \mathbf{e}_{ir} &= [\boldsymbol{\eta}_{ir}]_{\mathrm{sp}}. \end{aligned} \tag{45}$$

For the Lorentz model, the logarithmic map and parallel transport are given by Equations (29) and (30). For the Poincaré model, both points are mapped by Φ_c, the Lorentz operations are applied, and the shared spatial tangent coordinates are retained. For the Euclidean model, $\mathbf{e}_{ir} = \mathbf{z}_i - \mathbf{p}_r$.

The first-order consequent tangent and its manifold image are

$$\begin{aligned} \mathbf{u}_{ir} &= \mathrm{clip}_\tau(\mathbf{b}_r + \mathbf{W}_r \mathbf{e}_{ir}), \\ \mathbf{q}_{ir}^{\mathcal{M}} &= \exp^{\mathcal{M}}_{\mathbf{o}_{\mathcal{M}}}(\mathbf{u}_{ir}), \end{aligned} \tag{46}$$

where $\mathbf{b}_r \in \mathbb{R}^H$ and $\mathbf{W}_r \in \mathbb{R}^{H \times D}$. Setting $\mathbf{W}_r = \mathbf{0}$ gives the zero-order variant. Each consequent remains rule-specific through its base point $\mathbf{p}_r$ and affine parameters $(\mathbf{b}_r, \mathbf{W}_r)$, although aggregation uses a shared origin chart.

C.2 Fréchet Aggregation and Classification

The aggregate for sample i minimizes the weighted Fréchet objective

$$\begin{aligned} \mathcal{F}_i(\mathbf{m}) &= \frac{1}{2} \sum_{r=1}^{R} \overline{w}_{ir} d_c^{\mathcal{M}}(\mathbf{m}, \mathbf{q}_{ir}^{\mathcal{M}})^2, \\ \mathbf{m}_i^\star &= \operatorname*{argmin}_{\mathbf{m} \in \mathcal{M}} \mathcal{F}_i(\mathbf{m}). \end{aligned} \tag{47}$$

The Riemannian gradient is $\operatorname{grad} \mathcal{F}_i(\mathbf{m}) = -\sum_r \overline{w}_{ir} \log_{\mathbf{m}}(\mathbf{q}_{ir})$. The update below is therefore Riemannian gradient descent with step size η_{K}. The Fréchet minimizer is unique in hyperbolic space, but the finite Karcher iterations yield a differentiable approximation to the minimizer.

For Lorentz consequents, the normalized ambient initialization is

$$\begin{aligned} \mathbf{h}_i &= \sum_{r=1}^{R} \overline{w}_{ir} \mathbf{q}_{ir}^{\mathrm{L}}, \\ \mathbf{m}_i^{(0)} &= \frac{\mathbf{h}_i}{\sqrt{c\,(-\langle \mathbf{h}_i, \mathbf{h}_i \rangle_{\mathrm{L}})}}. \end{aligned} \tag{48}$$

This normalization maps $\mathbf{h}_i$ to $\mathcal{H}_c^H$. Because the positive firing weights sum to one, $\mathbf{h}_i$ remains in the future-directed timelike cone and the normalization selects the upper sheet. Poincaré consequents are first mapped through Φ_c, initialized by Equation (48), and mapped back after refinement. The Euclidean model uses the weighted arithmetic mean in Equation (39).

Starting from $\mathbf{m}_i^{(0)}$, HyperANFIS computes

$$\begin{aligned} \mathbf{d}_i^{(k)} &= \sum_{r=1}^{R} \overline{w}_{ir} \log^{\mathcal{M}}_{\mathbf{m}_i^{(k)}}(\mathbf{q}_{ir}^{\mathcal{M}}), \\ \mathbf{m}_i^{(k+1)} &= \exp^{\mathcal{M}}_{\mathbf{m}_i^{(k)}}(\eta_{\mathrm{K}} \mathbf{d}_i^{(k)}). \end{aligned} \tag{49}$$

The iterations terminate when $\|\mathbf{d}_i^{(k)}\| \le \epsilon_{\mathrm{K}}$ for every sample in the batch or after three steps. At a stationary point, $\sum_r \overline{w}_{ir} \log_{\mathbf{m}_i}(\mathbf{q}_{ir}) = \mathbf{0}$, which is the first-order condition of Equation (47).

For classification, class tangents $\mathbf{g}_k \in \mathbb{R}^H$ define

$$\begin{aligned} \mathbf{c}_k^{\mathcal{M}} &= \exp_{\mathbf{o}_{\mathcal{M}}}^{\mathcal{M}} \left[\mathrm{clip}_\tau(\mathbf{g}_k)\right], \\ s_{ik} &= -d_c^{\mathcal{M}}(\mathbf{m}_i^{\mathcal{M}}, \mathbf{c}_k^{\mathcal{M}})^2. \end{aligned} \tag{50}$$

C.3 Training Objective

For a mini-batch of size B, let $P_{ik} = \mathrm{softmax}_k(s_{ik})$ and let y_i be the class label. If n_k is the number of training samples in class k, the class weight is

$$\gamma_k = \frac{N}{K n_k}, \tag{51}$$

where N and K are the number of training samples and classes, respectively. The classification loss is

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{B} \sum_{i=1}^{B} \gamma_{y_i} \log P_{iy_i}. \tag{52}$$

Class weighting is applied only during training. Validation loss and macro-F1 are unweighted.

Define the batch-average rule usage

$$\pi_r = \frac{1}{B} \sum_{i=1}^{B} \overline{w}_{ir}. \tag{53}$$

The rule-balance term is the divergence of average rule usage from the uniform distribution:

$$\mathcal{L}_{\mathrm{bal}} = \sum_{r=1}^{R} \pi_r \log \left(\frac{\pi_r}{1/R} \right). \tag{54}$$

This term is minimized by uniform average usage and does not require uniform firing for each sample.

The rule-specialization term is the normalized firing entropy:

$$\mathcal{L}_{\mathrm{spec}} = -\frac{1}{B \log R} \sum_{i=1}^{B} \sum_{r=1}^{R} \overline{w}_{ir} \log \overline{w}_{ir}. \tag{55}$$

Minimizing this term encourages concentrated sample-level firing. Its coefficient follows the warm-up schedule

$$\kappa(t) = \min \left(\frac{t}{T_{\mathrm{w}}}, 1 \right), \tag{56}$$

which delays the full specialization penalty while rule prototypes establish coverage. The balance term discourages globally unused rules, whereas the specialization term discourages diffuse sample-level firing.

The prototype-separation term is

$$\begin{aligned} \mathcal{L}_{\mathrm{sep}} = \frac{1}{R(R-1)} \sum_{r \neq q} \\ \cdot \left[m_0 - d_c^{\mathcal{M}}(\mathbf{p}_r^{\mathcal{M}}, \mathbf{p}_q^{\mathcal{M}}) \right]_+^2. \end{aligned} \tag{57}$$

Only prototype pairs separated by less than m_0 contribute. The training objective is

$$\begin{aligned} \mathcal{L}(t) = \mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{bal}} \mathcal{L}_{\mathrm{bal}} \\ + \kappa(t) \lambda_{\mathrm{spec}} \mathcal{L}_{\mathrm{spec}} + \lambda_{\mathrm{sep}} \mathcal{L}_{\mathrm{sep}}. \end{aligned} \tag{58}$$

C.4 Optimization

The forward map is differentiable almost everywhere with respect to the tangent-space centers, scale logits, consequent parameters, and class prototypes. For either membership function, firing normalization satisfies

$$\frac{\partial \overline{w}_{ir}}{\partial \ell_{iq}} = \overline{w}_{ir} \left(\mathbf{1}[r = q] - \overline{w}_{iq} \right). \tag{59}$$

The log-membership derivatives are

$$\begin{aligned} \frac{\partial \ell^{\mathrm{G}}}{\partial \chi} &= -\chi, \\ \frac{\partial \ell^{\mathrm{B}}}{\partial \chi} &= -\frac{2b \operatorname{sign}(\chi) |\chi|^{2b-1}}{1 + |\chi|^{2b}}. \end{aligned} \tag{60}$$

The sigmoid parameterization in Equation (41) keeps each scale trainable within its prescribed interval. Radial clipping is the identity inside the tangent ball and removes the outward radial gradient component outside the ball. Its nondifferentiable boundary has measure zero; automatic differentiation uses the active branch.

The Karcher iterations in Equation (49) are unrolled in the computation graph. Gradients propagate through the logarithmic maps, weighted tangent sums, exponential updates, and ambient initialization. The discrete nearest-rule assignments are computed before training and are not differentiated. Early stopping and collapse detection do not modify the forward map.

The numerical clamps in Appendix E.2 are piecewise differentiable. An active clamp has zero gradient toward the invalid region, preventing updates across the Poincaré boundary or the domain boundary of arcosh.

Preprocessing and nearest-rule initialization precede gradient-based training. For each mini-batch, the model computes manifold embeddings, memberships, firing strengths, rule consequents, and the Karcher aggregate; evaluates Equation (58); and applies one optimizer update after backpropagation. Model-selection criteria are given in Appendix E.4.

D Theoretical Analysis

We establish the uniqueness of intrinsic aggregation and characterize the zero-curvature limit of the hyperbolic distance.

D.1 Uniqueness of Fréchet Aggregation

Proposition 1. For nonnegative weights $\overline{w}_{ir}$ summing to one, the Fréchet objective in Equation (47) has a unique minimizer in the Lorentz and Poincaré models.

Proof. Hyperbolic space is complete, simply connected, and has constant negative sectional curvature and is therefore a Hadamard manifold. On a Hadamard manifold, the weighted sum of squared distances is coercive and strictly geodesically convex. Completeness and coercivity give existence, and strict convexity gives uniqueness. The result holds in both coordinate models by the isometry Φ_c. In Euclidean space, the corresponding objective is a convex quadratic with minimizer $\sum_r \overline{w}_{ir} \mathbf{q}_{ir}$. □

HyperANFIS approximates this minimizer with at most three Karcher iterations and terminates earlier when the batch-wide tolerance is satisfied.

D.2 Euclidean Limit

Proposition 2. For fixed shared tangent vectors $\mathbf{v}$ and $\mathbf{a}$,

$$\lim_{c \downarrow 0} d_c^{\mathrm{L}} \left(\begin{matrix} \exp_{\mathbf{o}_{\mathrm{L}}}^{\mathrm{L}}(\mathbf{v}), \\ \exp_{\mathbf{o}_{\mathrm{L}}}^{\mathrm{L}}(\mathbf{a}) \end{matrix} \right) = \|\mathbf{v} - \mathbf{a}\|. \tag{61}$$

Proof. Expanding the hyperbolic sine and cosine terms in Equation (26) gives

$$\begin{aligned} -c \left\langle \begin{matrix} \exp^{\mathrm{L}}_{\mathbf{o}_{\mathrm{L}}}(\mathbf{v}), \\ \exp^{\mathrm{L}}_{\mathbf{o}_{\mathrm{L}}}(\mathbf{a}) \end{matrix} \right\rangle_{\mathrm{L}} \\ = 1 + \frac{c}{2}\|\mathbf{v}-\mathbf{a}\|^2 + O(c^2). \end{aligned} \tag{62}$$

Since $\operatorname{arcosh}(1+u) = \sqrt{2u} + O(u^{3/2})$ as $u \downarrow 0$, substitution into Equation (28) proves the claim. The same limit holds in Poincaré coordinates by isometry. □

Under the tangent bounds in Equation (21), the logarithmic maps, transported local coordinates, and Fréchet objective also converge to their Euclidean counterparts. The Euclidean variant is therefore the zero-curvature counterpart of the global radial-rule architecture, not a conventional ANFIS based on products of coordinate-wise memberships.

E Implementation Details

This section specifies preprocessing, initialization, numerical safeguards, and model selection.

E.1 Preprocessing and Initialization

All preprocessing statistics are estimated from the training partition. For raw feature x_{ij},

$$\widehat{x}_{ij} = \frac{x_{ij} - \nu_j}{\zeta_j}, \tag{63}$$

where ν_j and ζ_j are the training mean and standard deviation. Let $Q_{0.95}$ denote the empirical 0.95 quantile and let ρ be the target tangent radius. The fixed input scale is

$$\begin{aligned} \mathcal{R}_{\mathrm{tr}} &= \{\|\widehat{\mathbf{x}}_i\| : i \in \mathcal{D}_{\mathrm{tr}}\}, \\ r_{0.95} &= \max\left\{Q_{0.95}(\mathcal{R}_{\mathrm{tr}}), 10^{-8}\right\}, \\ s &= \frac{\rho}{r_{0.95}}. \end{aligned} \tag{64}$$

This normalization maps the empirical reference radius to ρ before radial clipping; τ remains a separate upper bound.

Initial center tangents are sampled from $\{\mathrm{clip}_\tau(s\widehat{\mathbf{x}}_i)\}$ without replacement when the training set contains at least R samples, and with replacement otherwise. For at most N_{init} randomly selected training samples, define

$$\begin{aligned} r^\star(i) &= \operatorname*{argmin}_{1\le r\le R} \delta^{\mathcal{M}}_{ir}, \\ d_i^{\mathrm{near}} &= \frac{1}{\sqrt{D}} \min_{1\le r\le R} \delta^{\mathcal{M}}_{ir}. \end{aligned} \tag{65}$$

The global scale estimate is the q_{init} quantile of $\{d_i^{\mathrm{near}}\}$. Rule r uses the same quantile of $\{\delta^{\mathcal{M}}_{ir}/\sqrt{D} : r^\star(i) = r\}$. Empty or invalid per-rule estimates are replaced by a positive global estimate. After multiplication by m_{init}, scales are clipped to $[\sigma_{\min}, \sigma_{\max}]$ and converted to the logit parameterization in Equation (41).

Consequent biases and class-prototype tangents are initialized from $\mathcal{N}(0, 0.05^2)$ and $\mathcal{N}(0, 0.15^2)$, respectively. Consequent matrices use Xavier-uniform initialization with gain 0.25. These initialized quantities remain trainable, whereas sample-to-rule assignments are used only during initialization.

E.2 Numerical Stability

Poincaré points are radially capped at $(1-\epsilon_{\mathrm{B}})/\sqrt{c}$ with $\epsilon_{\mathrm{B}} = 10^{-5}$. Lorentz projection preserves the spatial coordinates and recomputes the positive time coordinate as $\sqrt{c^{-1} + \|\mathbf{z}_{\mathrm{sp}}\|^2}$. The radial norm in the Poincaré projection uses the lower bound $\epsilon_{\mathrm{N}} = 10^{-12}$.

Denominators, squared norms, and nonnegative radicands use the lower bound ϵ_{N}. Arguments to arcosh use $1+\epsilon_{\mathrm{A}}$, where $\epsilon_{\mathrm{A}} = 10^{-7}$. This lower bound is an implementation safeguard; analytically, $\operatorname{arcosh}(1) = 0$. Explicit membership values use a lower bound of 10^{-12} for reporting, whereas firing strengths are computed from the unclamped log-memberships in Equation (44). Karcher iterates are reprojected after each step, and Poincaré aggregation is evaluated through the Lorentz isometry.

E.3 Feature-Space Interpretation

For a raw observation $\mathbf{x}$, antecedent evaluation follows

$$\begin{aligned} \mathbf{x} &\to \widehat{\mathbf{x}} = (\mathbf{x} - \boldsymbol{\nu}) \oslash \boldsymbol{\zeta} \\ &\to \mathbf{v} = \operatorname{clip}_{\tau}(s\widehat{\mathbf{x}}) \\ &\to \mathbf{z}^{\mathcal{M}} = \exp^{\mathcal{M}}_{\mathbf{o}_{\mathcal{M}}}(\mathbf{v}), \end{aligned} \tag{66}$$

where $\oslash$ denotes element-wise division. Reproducing this mapping requires the fitted preprocessing statistics and the scale s.

A rule center has shared tangent representative $\overline{\mathbf{a}}_r = \operatorname{clip}_{\tau}(\mathbf{a}_r)$. Its feature-space representative is

$$\widetilde{\mathbf{x}}_r = \boldsymbol{\nu} + \boldsymbol{\zeta} \odot \frac{\overline{\mathbf{a}}_r}{s}, \tag{67}$$

where $\odot$ denotes element-wise multiplication. This inverse gives a feature-space representative of the rule center. It is not globally unique, because inputs outside the tangent bound can map to the same clipped point.

E.4 Model Selection

Validation macro-F1 controls learning-rate scheduling, checkpointing, and early stopping. The test set is evaluated once with the checkpoint that attains the highest validation macro-F1. After a fixed grace period, a diagnostic terminates persistent single-class collapse without modifying Equation (58).