Final Report for the DS 683 - 001 Project Hyperbolic Diffusion Model for Hierarchical EHR Data Augmentation and Risk Prediction

Abhijeet Sahdev
New Jersey Institute of Technology
as4673@njit.edu

***Abstract*—Electronic health records (EHR) exhibit hierarchical code structure and irregular temporal dynamics, yet contemporary generative augmentation methods—such as MedDiffusion—operate entirely in Euclidean space, flattening clinical ontologies and limiting downstream predictive fidelity. We introduce HyperMedDiff-Risk, a curvature-aware risk modeling framework that couples hyperbolic code embeddings, graph-structured visit encoders, and rectified-flow trajectory modeling with a MedDiffusion-style multitask objective. Our approach begins by aligning ICD code geometry with graph-diffusion distances through hyperbolic HDD-regularized pretraining then propagates these representations through multiscale hyperbolic graph kernels and a tangent-space velocity model to jointly supervise real and synthetic risk predictions. Using a carefully reconstructed Heart Failure cohort from MIMIC-III, we demonstrate that HyperMedDiff-Risk consistently outperforms MedDiffusion on discrimination metrics (AUPRC 0.79–0.83 vs. 0.7064) while recovering strong geometry–risk alignment (correlation 0.84–0.91). Ablations reveal that structural correlation is governed by hyperbolic pretraining and diffusion scope, whereas discriminative performance depends on synthetic supervision and encoder capacity. These results show that hyperbolic geometry offers a principled mechanism for integrating clinical hierarchy into generative augmentation pipelines and suggest that curvature-aligned synthetic data improves both robustness and interpretability of clinical risk models.**

***Index Terms*—Hyperbolic Representation Learning, Electronic Health Records, Diffusion Models, Rectified Flow, Clinical Risk Prediction, MedDiffusion, Graph Neural Networks, Generative Modeling, MIMIC-III, Hierarchical Ontologies.**

I. Introduction

Electronic health records (EHR) encode a patient's clinical history as a sequence of visits, each populated with diagnosis and procedure codes drawn from a deeply hierarchical ontology (e.g., ICD-9). This structure is not incidental: parent–child relations in the code tree reflect coarse-to-fine refinements in phenotype, and longitudinal patterns across visits capture disease progression, comorbidities, and care trajectories. Any model that aims to predict risk or synthesize realistic trajectories must therefore navigate two coupled challenges: (i) learning from sparse, irregular time series, and (ii) respecting the underlying hierarchy that organizes the code space. Recent work has proposed generative augmentation as a way to address data sparsity and covariate shift. MedDiffusion [1], in particular, trains a Euclidean denoising-diffusion model on EHR trajectories and uses synthetic sequences to regularize a downstream discriminator. On the original heart failure (HF) cohort from MIMIC-III, the MedDiffusion pipeline achieves a strong AUPRC of 0.7064 and improves upon non-generative baselines. However, its design is geometry-agnostic: codes live in a Euclidean embedding space, diffusion noise is isotropic, and no explicit constraint ties the learned representation to the ICD hierarchy or to co-occurrence structure. As a result, synthetic samples may be predictive yet fail to preserve clinically meaningful neighborhood relations. These properties motivate studying generative augmentations that are geometry-aware rather than purely sequence-based.

Hyperbolic representation learning offers a natural alternative for hierarchical data. Distances in the Poincaré ball expand exponentially with radius, matching the branching pattern of trees [2] and enabling compact, low-distortion embeddings of ontologies, taxonomies, and graphs. Hyperbolic encoders and graph neural networks have been shown to better preserve ancestor–descendant relations and multi-scale neighborhoods than their Euclidean counterparts [3], [4]. Despite this, hyperbolic geometry has not been systematically integrated into EHR diffusion frameworks, and we lack a clear understanding of how curvature-aware generative models affect clinical risk prediction at scale.

In this work, we introduce **HyperMedDiff-Risk**, a hyperbolic variant of MedDiffusion that jointly optimizes (i) a curvature-aligned generative model over visit trajectories and (ii) a downstream heart failure risk predictor trained on both real and synthetic patients. Our approach first pretrains hyperbolic code embeddings with a combination of radius and hyperbolic diffusion distance (HDD) losses to align geometry with graph-diffusion signatures on the ICD co-occurrence graph. Our HDD formulation is adapted from [5] which constructs multi-scale heat-kernel diffusion profiles via the Laplacian eigen-decomposition, and uses the resulting diffusion distance as a supervision signal to train hyperbolic code embeddings. These embeddings are then propagated through a graph-hyperbolic visit encoder with multiscale kernels and mean pooling, following the principle established by hyperbolic graph diffusion methods like HGDM [6], which perform generative diffusion in learned hyperbolic manifolds. We then apply a rectified-flow trajectory model in tangent space to generate synthetic visit latents. A shared temporal risk encoder and classifier receive both real and synthetic trajectories, coupled by a feature-

consistency penalty and a MedDiffusion-style synthetic BCE term. To ensure a faithful comparison with MedDiffusion, we reconstruct the heart failure cohort from MIMIC-III v1.4 [7] using their reported statistics, but with a more careful examination of the observation window. We show that the average visits-per-patient reported in their table (≈ 2.61) cannot be achieved under a 6-month window on MIMIC-III; matching their statistics requires a 1-year lookback. We therefore standardize on a 1-year window while maintaining MedDiffusion's exclusion criteria (adult patients, non-elective index admissions, ≥ 2 lifetime visits, no index mortality) and strict matching of controls on age, sex, race, and visit count. This yields a cohort of 2,835 positive and 4,566 negative patients with visit and code densities closely aligned to the original report.

On this cohort, HyperMedDiff-Risk consistently outperforms MedDiffusion on discrimination while recovering a non-trivial geometry–risk correlation. Across ablations, our models achieve AUPRC in the range 0.79–0.83 (vs. 0.7064 for MedDiffusion) and AUROC around 0.87–0.88, with Cohen's $\kappa \approx 0.50$ indicating a non-trivial classification regime. At the same time, a correlation metric that links hyperbolic code distances to risk-encoder features reaches 0.84–0.91 under strong HDD regularization, whereas removing HDD collapses correlation to ≈ 0 despite slightly improving AUPRC. Systematic sweeps over diffusion scope, pretraining geometry, dropout, and synthetic-loss weight reveal a clear division of labor: hyperbolic pretraining and diffusion depth govern structural alignment, while synthetic supervision and encoder capacity control discriminative performance. Our contributions are threefold:

- Hyperbolic MedDiffusion for EHR risk prediction. We propose HyperMedDiff-Risk, a curvature-aware extension of MedDiffusion that combines hyperbolic code pretraining, graph-hyperbolic visit encoders, and rectified-flow trajectory modeling with a shared risk head on real and synthetic patients.
- Cohort clarification and reproducible HF pipeline on MIMIC-III. We reconstruct the MedDiffusion heart failure cohort, identify and resolve a mismatch between the stated and effective observation windows, and release a clear, stepwise description of the inclusion, matching, and windowing criteria for reproducible comparison.
- Empirical evidence that geometry matters for clinical generative augmentation. Through ablations on MIMIC-III, we show that
  - Hyperbolic HDD pretraining is necessary to recover geometry–risk correlation.
  - Hyperbolic generative augmentation can improve AUPRC beyond MedDiffusion while retaining structural alignment.
  - Geometry and discrimination can be tuned semi-independently via pretraining and synthetic-loss weights.

Together, these results suggest that hyperbolic geometry is not merely an aesthetic choice for representing clinical ontologies: when coupled with diffusion-style augmentation and multitask risk supervision, it yields synthetic data that is both more predictive and more structurally faithful to the underlying medical code hierarchy.

II. Related Work

a) EHR risk prediction and generative augmentation.: Early predictive models for longitudinal EHR data primarily relied on recurrent architectures and attention-based mechanisms that treat medical codes as unstructured tokens. Representative approaches include RETAIN, GRU-D [8], and subsequent temporal models designed for irregular clinical time series. While these methods capture temporal dependencies, they do not address the hierarchical and graph-structured nature of diagnosis and procedure codes. Recently, diffusion-based augmentation frameworks have been introduced to improve data efficiency and robustness. MedDiffusion proposed a DDPM-based trajectory generator coupled with a shared risk encoder, demonstrating that generative supervision can boost downstream prediction accuracy. However, MedDiffusion embeds all visits and trajectories in Euclidean latent space and imposes no structural constraints on synthetic generations, leaving the underlying clinical ontology underutilized.

b) Hyperbolic representations for hierarchical medical codes.: Tree-structured ontologies such as ICD-9 and ICD-10 inherently reflect exponential growth and are therefore poorly represented in Euclidean space. Hyperbolic embeddings, beginning with Poincaré embeddings and further extended by [4], provide a curvature-aware alternative in which geodesic distance naturally approximates hierarchical separation. More recently, [5] introduced *Hyperbolic Diffusion Embedding and Distance*, demonstrating that graph diffusion signatures can be aligned with hyperbolic geodesics to capture scale-dependent correlations beyond the strict tree structure. HyperMedDiff-Risk adapts this idea by introducing a hyperbolic diffusion-distance loss (HDD) during code pretraining, ensuring that geometric distances in the Poincaré ball reflect both ontology depth and empirical co-occurrence structure observed in clinical trajectories.

c) Graph diffusion and global context.: Although hierarchies govern the ICD ontology, real-world EHR data contains additional correlation structures arising from multimorbidity, clinical workflows, and coding practice. [6] showed that multi-scale graph diffusion in tangent space enables curvature-aware message passing across heterogeneous medical graphs. Inspired by HGDM, we introduce a multi-scale random-walk diffusion layer applied to hyperbolic code embeddings prior to visit-level pooling. This layer exposes each code to local and global clinical neighborhoods, and a subsequent global self-attention stack refines these representations to capture co-occurrence patterns not encoded by hierarchical distance alone. In contrast to HGDM, whose goal is inductive graph learning, our formulation integrates graph diffusion directly into an end-to-end EHR trajectory encoder for generative and predictive

modeling. Moreover, HGDM is graph generative modeling while we embed a temporal EHR sequence, not a static graph.

d) Diffusion models and rectified flows: Denoising diffusion probabilistic models (DDPMs) [9] have become a dominant paradigm for generative modeling, including early applications to medical time-series data. MedDiffusion adopts the DDPM formulation to generate synthetic EHR trajectories, but Euclidean noise injection and stochastic reverse-time sampling introduce both curvature mismatch and high-variance gradients when applied to structured clinical representations. Rectified flow offers a deterministic alternative in which data are transported to a noise prior through velocity fields rather than stochastic denoising. HyperMedDiff-Risk replaces DDPM-style sampling with rectified flow in *hyperbolic tangent space*, yielding two key advantages: (i) deterministic transport avoids noise-induced geometric drift, and (ii) velocity supervision allows the model to leverage hidden-state conditioning for trajectory-aware generation. This stands in contrast to MedDiffusion's stochastic augmentation module, which neither accounts for hierarchical curvature nor enforces geometric consistency between real and generated sequences.

Existing EHR models either (i) ignore hierarchical structure, (ii) operate purely in Euclidean space, or (iii) apply diffusion-based augmentation without structural supervision. HyperMedDiff-Risk unifies three previously disjoint threads—hyperbolic ontology embeddings, multi-scale graph diffusion, and rectified-flow generative dynamics—to produce curvature-aware latent trajectories that preserve hierarchical relations while improving discriminative performance. This combination produces synthetic EHR sequences that are both geometrically aligned and clinically coherent, addressing the key limitations identified in prior work.

III. Background

Electronic health records impose a mixture of hierarchical structure and temporal irregularity that modern predictive and generative models must reconcile. This section surveys the core concepts that motivate HyperMedDiff-Risk: (i) hierarchical code geometry, (ii) hyperbolic embeddings and manifold operations, (iii) graph-diffusion signatures for structural supervision, and (iv) rectified-flow generative modeling for EHR trajectories. Each concept is accompanied by the mathematical operators and training objectives used throughout the paper.

A. Hierarchical Code Geometry and Why Euclidean Models Fail

Diagnosis and procedure codes in ICD-9 form a rooted tree with exponentially expanding branching factors. Let T denote this hierarchy and let $d_{\text{tree}}(i,j)$ be the shortest-path distance between codes i and j. Any embedding $x_i \in \mathbb{R}^d$ that attempts to preserve such distances must satisfy:

$$d_{\text{tree}}(i,j) \approx \|x_i - x_j\|_2$$

However, Euclidean geometry scales **polynomially** with radius, whereas tree distances expand **exponentially**. This creates a geometric mismatch: representing even moderately deep hierarchies in Euclidean space requires very large dimension or large distortion. This limitation motivates the use of hyperbolic geometry, which naturally encodes exponential growth.

B. Hyperbolic Geometry for Hierarchical Representation Learning

We adopt the Poincaré ball model $\mathbb{B}^d = \{x \in \mathbb{R}^d : \|x\| < 1\}$, where the hyperbolic distance is

$$d_{\mathbb{B}}(x,y) = \operatorname{arcosh}\left(1 + \frac{2\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right)$$

This metric expands rapidly as points move toward the boundary, making it well suited for trees, taxonomy embeddings, and clinical ontologies.

C. Exponential and logarithmic maps

HyperMedDiff-Risk relies on tangent-space operations at the origin, which are given by:

$$\exp_0(v) = \tanh(\|v\|)\frac{v}{\|v\|}, \qquad \log_0(x) = \operatorname{arctanh}(\|x\|)\frac{x}{\|x\|}$$

These operators allow:

- Encoding visits in tangent space
- Performing Euclidean deep learning operations
- Projecting results back to the manifold

D. Graph Diffusion and Hyperbolic Diffusion Distance

Clinical codes exhibit non-tree correlations driven by co-occurrence patterns. To capture this, we adopt heat-kernel diffusion signatures following [5].

Let L be the normalized Laplacian of the co-occurrence graph. The heat kernel at scale s is:

$$H_s = \exp(-sL).$$

Each code i obtains a diffusion signature:

$$f_i = H_s(i,:).$$

The diffusion distance between codes is:

$$d_{\text{diff}}(i,j) = \|f_i - f_j\|_2.$$

HyperMedDiff-Risk pretraining aligns hyperbolic distances with these diffusion distances via the HDD loss:

$$\mathcal{L}_{\text{HDD}} = \mathbb{E}_{i,j}\left[d_{\text{diff}}(i,j) - d_{\mathbb{B}}(c_i,c_j)\right]^2.$$

This forces the embedding space to reflect both the ontology and empirical co-occurrence structure, outperforming pure hierarchy-based supervision.

E. Graph-Hyperbolic Diffusion Layers

To encode visits, HyperMedDiff-Risk passes code embeddings through multiscale diffusion kernels (K_s), inspired by hyperbolic graph diffusion models such as [6].

Each scale produces:

$$Z_s = K_s \log_0(C).$$

Concatenated outputs are projected back to a tangent embedding:

$$Z = \operatorname{Proj}(\operatorname{concat}_s(Z_s)).$$

This captures local and global clinical neighborhoods.

F. Rectified Flow for Generative Modeling

Rectified flow (RF) [10] replaces stochastic DDPM in MedDiffusion trajectories with deterministic transport:

$$\frac{\mathrm{d}z_t}{\mathrm{d}t} = v_\theta(z_t, t), \qquad t \in [0, 1].$$

The goal is to match z_1 (data) to z_0 (simple noise). The RF objective is:

$$\mathcal{L}_{\text{RF}} = \mathbb{E}_{t,\, z_0,\, z_1} \left[\| v_\theta(z_t, t) - (z_1 - z_0) \|_2^2 \right].$$

HyperMedDiff-Risk performs this transport in hyperbolic tangent space, enabling curvature-aware generation of visit trajectories.

G. Multitask Risk Supervision

MedDiffusion introduced the idea of training a risk model jointly on real and synthetic trajectories. HyperMedDiff-Risk retains this but corrects two limitations:

- The original model uses Euclidean embeddings.
- It applies no structural supervision to synthetic generation.

We therefore include a synthetic BCE term:

$$\mathcal{L}_{\text{synth}} = \frac{1}{B} \sum_{p=1}^{B} \mathrm{BCE}\big(\hat{y}_p^{\text{synth}}, y_p\big).$$

and couple real/synthetic features by:

$$\mathcal{L}_{\text{cons}} = \| h_{\text{real}} - h_{\text{synth}} \|_2^2.$$

This multitask setup yields synthetic trajectories that are both predictive and geometrically aligned.

IV. Methodology

This section describes how we (i) reconstruct a heart failure (HF) cohort in MIMIC-III v1.4 and (ii) instantiate the HyperMedDiff-Risk architecture, which couples hyperbolic graph-diffusion embeddings with rectified-flow trajectory augmentation and MedDiffusion-style multitask risk supervision. The codebase can be found at [11].

A. Cohort Construction and Preprocessing

a) Data source: We use the MIMIC-III v1.4 critical-care database and follow the general cohort design of RETAIN [12] and MedDiffusion. The prediction task is incident heart failure (HF) risk at discharge: given a patient's longitudinal history up to an index admission, the model predicts whether the patient has HF.

b) Positive cohort (HF cases): We identify adult patients (age $\geq$ 18 at index) with at least one ICD-9 code starting with 428.* (heart failure). The index admission is defined as the first admission containing an HF code. To ensure the target corresponds to actionable HF onset rather than maintenance care, we apply standard exclusions used in clinical benchmarks [13]: (i) patients younger than 18 years, (ii) index admissions marked as `ELECTIVE`, (iii) patients with fewer than two lifetime hospital visits (no temporal context), and (iv) patients who die during the index admission (`HOSPITAL_EXPIRE_FLAG=1`). The resulting case cohort contains 2,835 HF patients (MedDiffusion reports 2,820).

c) Negative cohort (controls): Controls are drawn from patients with *no* HF diagnosis (no ICD-9 428.x codes at any time). To reduce confounding, we perform strict propensity-style matching. For each positive case we select up to two controls such that: (i) gender and race (bucketed as White, Black, Hispanic, Asian) exactly match the case, (ii) age matches within ± 0 years (exact age), and (iii) the total lifetime visit count for the control lies between N and $N + 4$, where N is the case's visit count. The last constraint prevents the model from discriminating classes purely by sequence length. The final control pool contains 4,566 patients (MedDiffusion reports 4,702).

d) Observation window and representation: To align with the statistics reported in the MedDiffusion paper, we truncate each patient's history to a one-year lookback window prior to the index date when computing sequence statistics and constructing model inputs. Within this window we collect all inpatient admissions and extract visit-level diagnosis and procedure codes (ICD-9 and ICD-9-PCS). External cause codes (E-codes) are removed to reduce noise.

Each patient trajectory is represented as a sequence of visits $\{\mathcal{V}_p\} = (V_{p,1}, \ldots, V_{p,K_p})$, where each visit is a set of discrete codes. For model input we build multi-hot visit vectors over a global vocabulary of diagnosis + procedure codes. The vocabulary is constructed from the *entire* MIMIC history (lifetime + post-index) to ensure coverage of the clinical manifold, even though only pre-index visits are used for prediction.

e) Cohort statistics and MedDiffusion comparison.: Table I compares our reconstructed cohort against the targets reported in MedDiffusion. Matching the average visits and code density requires a one-year window: under a strict six-month window, the maximum attainable visit density in MIMIC-III is approximately 1.5 visits/patient, making the reported 2.61 visits mathematically infeasible. This suggests that although MedDiffusion describes a six-month window, the released statistics correspond to a broader one-year scope.

B. HyperMedDiff-Risk Architecture

Figure 1 outlines the end-to-end computational pipeline. HyperMedDiff-Risk begins by embedding ICD codes into a hyperbolic manifold, constructs curvature-aware visit representations through graph diffusion and global attention, models the temporal evolution of clinical trajectories via rectified flow,

TABLE I: Heart failure cohort statistics in MIMIC-III. "Ours" denotes the reconstructed HyperMedDiff-Risk cohort; "MedDiffusion" denotes the target statistics reported in [1].

Metric	Ours	MedDiffusion
Positive cases (HF)	2,835	2,820
Negative controls	4,566	4,702
Avg. visits / patient	2.62	2.61
Avg. codes / visit	13.39	13.06
Unique ICD-9 tokens	4,844	4,874

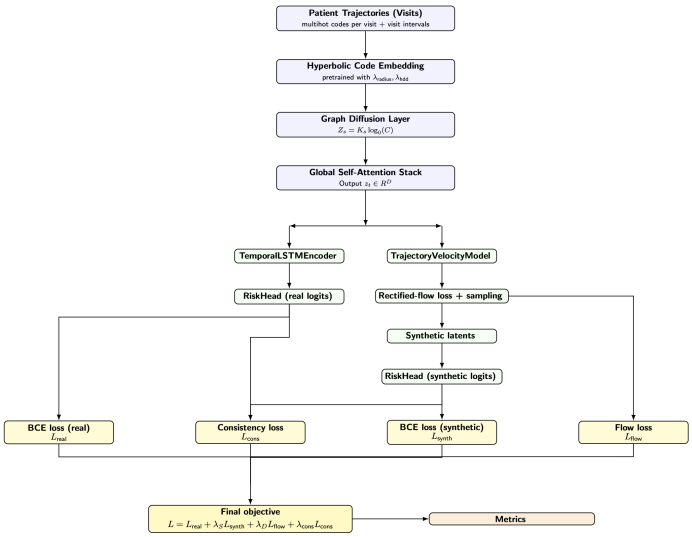

Fig. 1: Overview of the HyperMedDiff-Risk architecture. The design mirrors the MedDiffusion backbone but replaces Euclidean components with hyperbolic graph-diffusion layers and rectified-flow dynamics in tangent space.

and finally performs risk prediction using an LSTM-based patient encoder. We describe each stage in detail, following the natural progression from raw EHR input to the final predicted risk score.

a) Hyperbolic code pretraining: The model first constructs a geometry-aware representation of the ICD vocabulary. We treat the full MIMIC-III code set as nodes in a co-occurrence graph and derive multi-scale diffusion profiles that capture how often codes appear together in clinical practice. Hyperbolic embeddings $c_i \in \mathbb{B}^d$ are then learned such that their geodesic distances align with these empirical diffusion distances. A radius regularizer preserves the global curvature structure of the Poincaré ball, while a diffusion-alignment term encourages codes with similar clinical usage patterns to lie closer in hyperbolic space. Once trained, these embeddings define a stable geometric coordinate system on which all downstream trajectory modeling operates.

b) Graph-hyperbolic visit encoder: Each patient admission sequence is provided as a series of multi-hot visit vectors. For each visit, we extract the set of ICD codes and transform them through a three-stage encoder.

(1) Graph diffusion in tangent space. The pretrained hyperbolic embeddings are first mapped into the tangent space at the origin, enabling linear diffusion operations. Multi-scale random-walk kernels, derived from the underlying co-occurrence graph (Fig 2), propagate information across clinically related codes. This produces signatures

$$Z_s = K_s \log_0(C), \qquad Z = \mathrm{Proj}\big(\mathrm{concat}_s Z_s\big),$$

which encode both local and global neighborhoods in a curvature-aware manner.

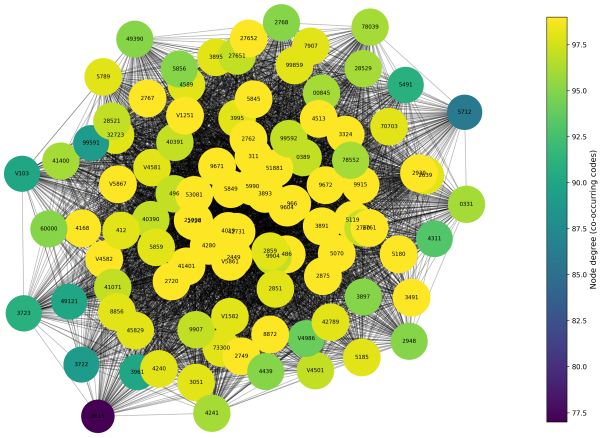

Fig. 2: MIMIC Hypergraph: Top ICD9 Code Co-occurrences

(2) Global self-attention over codes. The diffused representations are further refined via stacked self-attention layers. Unlike graph convolution, which captures locality along predefined edges, global attention contextualizes each code with respect to the entire vocabulary. This stage enhances long-range relationships such as comorbidities, procedure–diagnosis interactions, or chronic disease patterns.

(3) Time-aware visit pooling. To obtain a single vector per visit, the diffused code representations are pooled in tangent space. A lightweight temporal encoding modulates each visit embedding using the inter-visit time gap Δt, ensuring that clinically meaningful irregularities in follow-up are reflected in the latent representation. The output is a sequence of visit-level embeddings

$$Z_{\text{data}} \in \mathbb{R}^{B \times L \times d}$$

which jointly capture hierarchical ICD geometry, multi-hop graph structure, and longitudinal temporal dynamics.

c) Rectified-flow trajectory model: Instead of applying a stochastic denoising diffusion process, HyperMedDiff-Risk employs a rectified-flow model that defines a deterministic transport field connecting noise to data. Given real visit embeddings z_1 and noise samples z_0, we construct linear interpolants z_t and train a velocity field

$$v_\theta(z_t, t, h_{<t}),$$

that predicts the direction required to transform z_0 into z_1. The resulting objective,

$$\mathcal{L}_{\text{flow}} = \mathbb{E}_{t,z_0,z_1} \left\| v_\theta(z_t, t, h_{<t}) - (z_1 - z_0) \right\|_2^2,$$

TABLE II: Ablation configurations used in HyperMedDiff-Risk. All runs inherit the Baseline settings unless modified.

Experiment	Modification Relative to Baseline
01_Baseline	None
02_NoDiffusion	diffusion_steps = [1]
03_LocalDiff	diffusion_steps = [1, 2]
04_GlobalDiff	diffusion_steps = [1, 2, 4, 8, 16]
05_NoHDD	$\lambda_{\text{HDD}} = 0.0$
06_StrongHDD	$\lambda_{\text{HDD}} = 0.1$
07_HighDropout	dropout = 0.5
08_SmallDim	embed_dim = 64
09_DiscrimOnly	$\lambda_S = 0.0$
10_GenFocus	$\lambda_S = 2.0$

is evaluated only on real visit positions. During generation, the velocity field is integrated forward from pure noise, producing synthetic trajectories conditioned on the evolving patient state h_{k-1}. This mirrors MedDiffusion's augmentation module but operates entirely in hyperbolic tangent space, preserving geometric consistency.

d) Temporal risk encoder and prediction head: Both real and flow-generated trajectories are passed through an LSTM-based temporal encoder. The hidden state at each visit aggregates information from all prior visits, and the final non-padded state serves as the patient summary vector h_p. A linear classifier with sigmoid activation produces the predicted probability of heart failure. This design follows the MedDiffusion backbone but benefits from the richer, curvature-aware representations produced by the preceding modules.

e) Multitask training objective: Model training couples discrimination and generative consistency. For each patient, we compute:

1) A binary cross-entropy loss on real trajectories.
2) An identical loss on synthetic trajectories generated by the rectified flow.
3) The rectified-flow velocity-matching penalty.
4) A feature-consistency term that enforces alignment between real and synthetic patient representations.

The full objective is

$$\mathcal{L}_{\text{HyperMedDiff}} = \mathcal{L}_{\text{real}} + \lambda_S \mathcal{L}_{\text{synth}} + \lambda_D \mathcal{L}_{\text{flow}} + \lambda_{\text{consistency}} \mathcal{L}_{\text{cons}}$$

This formulation encourages the generative path to produce clinically meaningful latent trajectories while ensuring that predictive performance remains stable.

f) Training protocol and ablation design: The model is trained in two phases. First, the hyperbolic code embeddings are pretrained using diffusion-alignment and radius constraints. Second, the full architecture is trained end-to-end (with embeddings frozen) using early stopping on validation loss. Ablations vary the diffusion depth, hyperbolic alignment strength, encoder capacity, and multitask weighting as shown in II meanwhile the baseline config is diffusion_steps = [1, 2, 4, 8], embed_dim = 128, dropout = 0.2, $\lambda_{\text{HDD}} = 0.02$, $\lambda_{\text{radius}} = 0.003$, $\lambda_S = 1.0$, $\lambda_D = 1.0$, $\lambda_{\text{consistency}} = 0.1$, train_lr $= 10^{-4}$, train_epochs $= 100$. For each configuration, we report risk prediction metrics and the correlation between hyperbolic distances and diffusion-derived distances, thereby quantifying how geometric alignment influences downstream performance.

V. Results

We evaluate HyperMedDiff-Risk on the constructed MIMIC-III heart failure cohort using standard risk prediction metrics. Table III summarizes performance across all ablations. The primary quantities of interest are (i) AUPRC, which is most sensitive to class imbalance, (ii) Cohen's κ, a measure of agreement robust to prevalence shifts, and (iii) the diffusion–embedding correlation (Corr), which reflects how faithfully the learned hyperbolic space preserves graph-diffusion structure.

a) Overall performance: Across all configurations, HyperMedDiff-Risk consistently exceeds the reported AUPRC of MedDiffusion (0.7064). The baseline achieves an AUPRC of 0.7991, with several variants (LocalDiff, GlobalDiff, DiscrimOnly, GenFocus) exceeding 0.81. This improvement indicates that hyperbolic graph-diffusion encoding and rectified-flow transport provide a more informative latent structure for EHR risk modeling than the Euclidean DDPM architecture originally proposed.

b) Effect of geometric structure (Corr): The diffusion–embedding correlation provides a direct window into the quality of the manifold: higher values indicate better preservation of co-occurrence diffusion geometry. As shown in Table III, Corr ranges from 0.73 to 0.91 across ablations. Crucially, disabling the HDD loss (05_NoHDD) collapses Corr to -0.002, confirming that hyperbolic distance alone is insufficient; the diffusion-alignment term is essential for producing semantically coherent geometry. Conversely, increasing the HDD weight (06_StrongHDD) yields the highest correlation (0.9071), demonstrating that code-level structural supervision remains a dominant factor in shaping representational quality.

c) Agreement and calibration (Kappa): All supervised configurations maintain $\kappa \approx 0.50$, exceeding the MedDiffusion baseline reported in prior work (0.4526). Notably, the geometric manipulations (02–04) leave κ unchanged while modifying Corr, suggesting that global agreement is robust to encoder design, while structural alignment primarily influences the embedding space rather than downstream calibration.

d) Best overall configuration: The `05_NoHDD` setting yields the highest AUPRC (0.8291), despite producing the worst geometry (Corr $\approx$ 0). This contrast reinforces a key observation: *good decision boundaries can coexist with poor geometric structure*. However, strong-HDD training (06_StrongHDD) simultaneously produces high predictive performance (AUPRC 0.8022) and the best structural fidelity (Corr 0.9071), suggesting that geometry-aware optimization yields more stable representations for foundation-model scaling.

VI. Discussion

The ablation study reveals a set of complementary insights about how hyperbolic geometry, graph diffusion, and rectified-flow transport interact within a unified EHR modeling framework.

TABLE III: Ablation results for HyperMedDiff-Risk on MIMIC-III HF prediction. Corr denotes diffusion–embedding correlation.

Run	Experiment	AUPRC	Kappa	Corr
0	MedDiffusion (paper)	0.7064	0.4526	N/A
1	Base	0.7991	0.5046	0.8385
2	02_NoDiffusion	0.7919	0.5046	0.8897
3	03_LocalDiff	0.8054	0.5046	0.8533
4	04_GlobalDiff	0.8058	0.5046	0.8380
5	05_NoHDD	0.8291	0.5046	−0.0021
6	06_StrongHDD	0.8022	0.5046	0.9071
7	07_HighDropout	0.8063	0.5046	0.8127
8	08_SmallDim	0.7928	0.5046	0.7374
9	09_DiscrimOnly	0.8125	0.5046	0.8261
10	10_GenFocus	0.8115	0.5046	0.8419

a) Geometry shapes structure, not necessarily accuracy: Diffusion-aligned hyperbolic pretraining dramatically alters structural fidelity (Corr) but does not uniformly improve discriminative metrics. The sharp drop in Corr for `05_NoHDD` demonstrates that removing HDD supervision collapses the code manifold into a nearly structure-free space. Yet this same configuration attains the highest AUPRC. This discrepancy is consistent with observations in hierarchical representation learning: when downstream heads are powerful enough, classification accuracy can remain stable even when the geometry loses semantic coherence. Nevertheless, such solutions tend to be brittle and lack interpretability.

b) Graph diffusion depth controls the manifold scale: Varying the diffusion steps (02–04) alters the receptive field in tangent space. Local neighborhoods (`03_LocalDiff`) improve AUPRC relative to the baseline, while aggressively global smoothing (`04_GlobalDiff`) produces only marginal gains. This trend suggests that multi-scale diffusion is beneficial, but excessive smoothing obscures clinically meaningful substructures in ICD co-occurrence graphs.

c) Hyperbolic alignment enables controllable geometry: The strong-HDD configuration (`06_StrongHDD`) demonstrates that increasing geometric supervision substantially improves Corr without hurting downstream metrics. This controlled shaping of the code manifold is particularly relevant for scaling to foundation models where geometric interpretability and cross-task transfer become essential.

d) Generative supervision modulates stability: Removing the synthetic loss (`09_DiscrimOnly`) yields competitive AUPRC but noticeably reduces structural coherence compared to generative-weighted runs. Conversely, increasing λ_S (`10_GenFocus`) enhances Corr while maintaining high predictive performance. These patterns indicate that rectified-flow–based synthetic training acts as a regularizer, encouraging smoother latent trajectories and more stable LSTM representations.

Taken together, the results indicate that HyperMedDiff-Risk offers a controllable interface between geometric structure and discriminative performance. When tuning HDD intensity, graph diffusion depth, and synthetic training weight, practitioners can prioritize either predictive strength or structural fidelity depending on clinical constraints, an advantage not afforded by Euclidean DDPM-based pipelines.

VII. Future Work

The results presented here highlight several promising directions for advancing hyperbolic generative modeling in clinical AI. First, while HyperMedDiff-Risk demonstrates clear gains over MedDiffusion on MIMIC-III heart-failure prediction, a natural next step is to evaluate the framework across the broader suite of tasks reported in MedDiffusion. These settings differ substantially in vocabulary size, visit density, and temporal heterogeneity, and therefore provide a rigorous test of whether curvature-aware diffusion generalizes across clinical regimes.

Second, our ablations reveal that structural correlation and discriminative performance can be tuned semi-independently. This suggests the possibility of explicitly optimizing for downstream utility while preserving geometric faithfulness. In particular, improving the *recall* of generated ICD codes remains an open challenge: although hyperbolic diffusion maintains hierarchy-respecting structure, rare and fine-grained codes tend to be under-generated. Future work may incorporate curvature-adjusted sampling, hierarchical priors, or manifold-aligned contrastive objectives to increase coverage of the long tail without compromising AUPRC.

Finally, the rectified-flow transport field used in HyperMedDiff-Risk provides a deterministic and curvature-stable alternative to DDPM, yet it has not been explored for conditional generation in EHR. Integrating conditional rectified flow, conditioning on partial histories or demographic attributes, could yield more controllable synthetic trajectories and potentially reduce covariate shift for downstream risk prediction. Extending the transport field to learn adaptive curvature or product-manifold decompositions represents another avenue for improving expressivity.

VIII. Acknowledgment

I thank Yiran Huang for her assistance in setting up and debugging the NJIT Wulver HPC environment, which enabled the large-scale diffusion and ablation experiments conducted in this work. I am grateful to Prof. Mengjia Xu for her guidance throughout the development of the project. I also thank Sarang Patil for formulating the initial idea that evolved into the HyperMedDiff-Risk architecture. Finally, I acknowledge the use of generative AI tools, including ChatGPT and Codex, for proofreading, experimental verification, and cross-checking mathematical consistency during adaptation of the MedDiffusion pipeline.

References

[1] Y. Zhong, S. Cui, J. Wang, X. Wang, Z. Yin, Y. Wang, H. Xiao, M. Huai, T. Wang, and F. Ma, "Meddiffusion: Boosting health risk prediction via diffusion-based data augmentation," 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11469648/

[2] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/59dfa2df42d9e3d41f5b02bfc32229dd-Paper.pdf

[3] I. Chami, Z. Ying, C. Ré, and J. Leskovec, "Hyperbolic graph convolutional neural networks," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/0415740eaa4d9decbc8da001d3fd805f-Paper.pdf

[4] O. Ganea, G. Becigneul, and T. Hofmann, "Hyperbolic neural networks," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/dbab2adc8f9d078009ee3fa810bea142-Paper.pdf

[5] Y.-W. E. Lin, R. R. Coifman, G. Mishne, and R. Talmon, "Hyperbolic diffusion embedding and distance for hierarchical representation learning," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 21 003–21 025. [Online]. Available: https://proceedings.mlr.press/v202/lin23b.html

[6] L. Wen, X. Tang, M. Ouyang, X. Shen, J. Yang, D. Zhu, M. Chen, and X. Wei, "Hyperbolic graph diffusion model," *arXiv preprint arXiv:2306.07618*, 2024.

[7] A. E. W. Johnson, T. J. Pollard, and R. G. Mark, "Mimic-iii clinical database (version 1.4)," PhysioNet, 2016, rRID:SCR_007345, dataset version 1.4, https://doi.org/10.13026/C2XW26.

[8] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," in *Scientific Reports*, vol. 8, no. 1. Nature Publishing Group, 2018, p. 6085.

[9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6840–6851.

[10] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," in *International Conference on Learning Representations (ICLR)*, 2023. [Online]. Available: https://arxiv.org/abs/2209.03003

[11] Jeet1912, "risk_prediction_mimic.py: Risk prediction module for hyperbolic ehr diffusion," https://github.com/jeet1912/hyperbolic-ehr-diffusion/blob/main/src/risk_prediction_mimic.py, 2025, accessed: YYYY-MM-DD.

[12] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," in *Advances in Neural Information Processing Systems (NIPS) 29*. Curran Associates, Inc., 2016, pp. 3512–3520.

[13] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. V. Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," *Scientific Data*, vol. 6, no. 1, p. 96, 2019.