HNMP-EHR: ONTOLOGY-ALIGNED REPRESENTATION LEARNING ON ELECTRONIC HEALTH RECORDS FOR CONCURRENT PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

The growing availability of electronic health records (EHRs) presents ever more opportunities for data-driven applications in healthcare. However, these records often suffer from fragmentation and structural inconsistencies, where key contextual elements are frequently missing. In this work, we present a system that enhances clinical predictions based on representation learning through ontology alignment. To address data gaps, we introduce a dual-space strategy that guides the learning process over heterogeneous entities. To address structural conformity, we introduce a hyperbolic message passing algorithm that allows the integration of hierarchical information. Our experiments, focused on cardiovascular disease (CVD) patients, demonstrate that the system improves performance across multiple tasks, including lab test, diagnosis, medication, and mortality prediction, performed concurrently. The results highlight the potential of ontology-aligned representation learning for high-precision and multi-purpose decision-making in healthcare systems.

1 INTRODUCTION

The amount of stored data as electronic health records (EHR) has grown significantly in recent years, now including an immense quantity of interactions, events and interconnected information. As such, data integration can play a transformative role in the coming years for model development and decision making in health care. The integration processes however remain largely (semi) manual, introducing costly clinical errors—such as prescription mistakes and procedural oversights— stemming from incomplete knowledge, ambiguous records, or gaps in communication Franklin et al. (2005); Velo & Minuz (2009); Ross et al. (2009). With the growing availability of clinical datasets, predictive models have the potential to mitigate such risks, optimize clinical procedures, and ultimately improve patient treatment. This motivation has led to advances in research across many predictive tasks concerning diagnosis Choi et al. (2017b); Sun et al. (2021); Wu et al. (2021); Choi et al. (2020); Ma et al. (2018), medication Shang et al. (2019); Yang et al. (2021); Mishra & Shridevi (2024), and medical procedures Fouladvand et al. (2023); Mao et al. (2022), focusing on single and secondary-care use cases. Their adoption for primary-care and multi-purpose systems is however constrained by the requirements for high precision, incorporation of expert knowledge, as well as technological challenges due to heterogeneous, multi-source data Ferraz et al. (2025).

Graph representation learning offers a powerful approach for modeling the complex structure of EHR in data-driven health care Zhang et al. (2021). In particular, graph neural networks are able to capture local and global dependencies across diverse EHR sources, including patient histories and medical codes, through non-linear neural operators and message passing. These techniques can enhance predictive performance and patient centric decision making in many clinical tasks ranging from disease prediction to drug recommendation. In practical settings however, the quality of graph-based representations is undermined by data incompleteness and inconsistencies as a source of noise originating in legacy systems Mohamed et al. (2021).

The presence of noise remains a significant obstacle in graph-based modeling, very often leading to ambiguous representations. To address this issue, domain-specific ontologies and coding systems can be leveraged as a source structured medical knowledge to help reduce inconsistencies. In Hao et al. (2021) for instance, a graph learning approach is proposed that maps data sources to a domain specific reference ontology in order to resolve ambiguous representations. Learning EHR representations through alignment for single down stream tasks is further investigated in Mehryar (2025). Existing approaches remain largely task-specific, and their applicability across diverse clinical objectives is limited.

In order to enable multi-purpose and high-precision applications, in this paper we proposes a framework that enhances concurrent performance by integrating ontological knowledge with patient representations through hy-

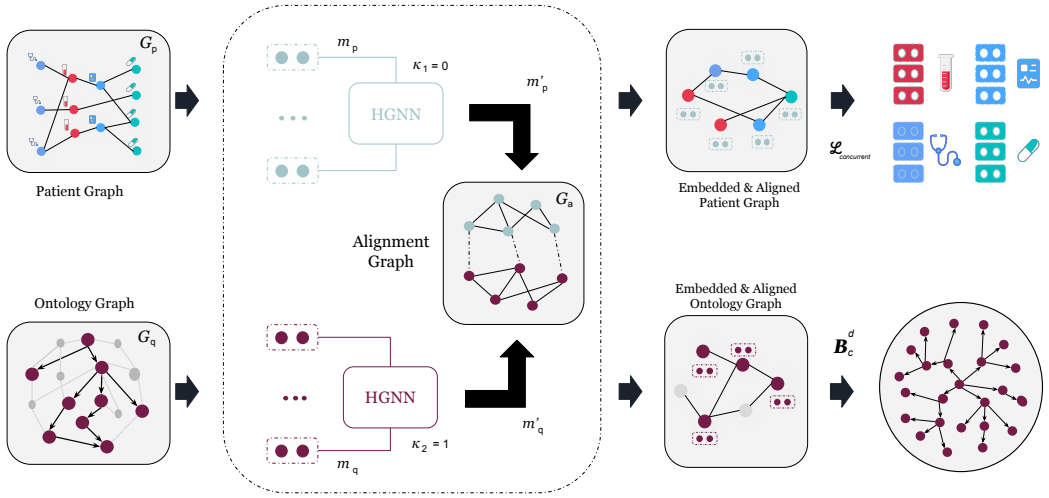


Figure 1: The embedding and alignment framework with dual-space hyperbolic neural message passing for ontological representation learning and concurrent prediction over clinical records.

perbolic message passing. To this end, a set of intra- and inter-graph message expressions are derived to facilitate the transfer of task-relevant information over variably structured spaces Iyer et al. (2022). We then introduce a multi-objective function that combines task-specific and hierarchical learning for concurrent predictions. While single and general purpose methods over real-world EHR exist, to the best of our knowledge learning adaptive representations via ontology-aligned hyperbolic message passing for concurrent applications is a novel approach.

2 METHOD

In this section, we present the proposed hybrid framework for EHR representation through alignment with a reference ontology, as illustrated in Figure 1. We conceptualize EHR data as an structureless graph, focusing on medical codes as nodes and representing their associations via edges (e.g. a patient has diagnosis Hypertension). Conversely, the reference ontology is provided as a directed acyclic graph that organizes medical concepts via parent-child relationships (e.g. Disease to Cardiovascular Disease to Hypertension). The representations for each graph is learned using a different embedding model that depends on the topological and structural information it contains. In particular, we consider one hyperbolic space to learn the representations over the EHR data with curvature κ_1 and a different hyperbolic space to learn the ontological representations with curvature κ_2. The dual representations are aligned using neural messages passed between a set of anchoring nodes.

More formally, let $\mathcal{O}_p$ denote medical observations, $\mathcal{D}_p$ denote diagnoses, and $\mathcal{M}_p$ denote medications over data from P patients. The set of all encoded EHR organized across visits is given by $\mathcal{C}_p = \mathcal{O}_p \cup \mathcal{D}_p \cup \mathcal{M}_p$. For a given patient p, the t'th visit v_p^t contains a subset of codes from $\mathcal{C}_p$ (i.e. $v_p^t \in \mathcal{C}_p$). Conversely, concepts from a reference ontology $\mathcal{C}_q$ are organized as a collection of ontology observations $\mathcal{O}_q$, ontology diagnoses $\mathcal{D}_q$, and ontology medications $\mathcal{M}_q$, i.e. $\mathcal{C}_q = \mathcal{O}_q \cup \mathcal{D}_q \cup \mathcal{M}_q$. The correspondence between patient and ontology sets is established through a non-empty set $\mathcal{A}$ of anchors, such that for any pair $(p, q) \in \mathcal{A}$, the node p (corresponding to an observation, diagnosis, or medication from the patient data) matches the node q (corresponding to an observation, diagnosis, or medication in the ontology). In particular, the set of anchor pairs for observation $\mathcal{A}_O$, diagnosis $\mathcal{A}_D$, and medication $\mathcal{A}_M$ are provided as $\mathcal{A} = \mathcal{A}_O \cup \mathcal{A}_D \cup \mathcal{A}_M$.

The latent representations are learned by applying neural message passing in two settings: along paths inside visits and ontology subsets (intra), as well as, across edges that connect anchor nodes (inter). Specifically, a collection of edges $\mathcal{R}_p$ captures the relationships among entities within a visit v_p^t (e.g. visit to diagnosis, procedure, or medication). In parallel, a collection of edges $\mathcal{R}_q$ is taken to encode semantic relations within the ontology, spanning observations $\mathcal{O}_q$, findings $\mathcal{D}_q$, and substances $\mathcal{M}_q$. In addition, anchor representations are aligned through dedicated message computations. These intra-visit, intra-ontology, and cross-anchor message-passing operations form the basis of the proposed representation learning framework. In particular, in this work we consider a hyperbolic space with $\kappa_1 = 0$ corresponding to the Euclidean space for message passing over the patient graph (structure-less) and a hyperbolic space with $\kappa_2 = 1$ corresponding to the Poincare model to represent the ontology graph (hierarchical), as described in the following sections.

2.1 PATIENT GRAPH MESSAGE COMPUTATIONS

The patient graph $\mathcal{G}_p$ includes structure-less information related to patient visits encoded by $\mathcal{C}_p$. The representations for patient visits are learned using hyperbolic neural message passing in an L layer graph neural network. For each node p_i from a visit v_p^t composed of heterogeneous neighbors $\mathcal{N}_i$, type-aware messages propagated along edges $\mathcal{R}_p$ guide the learning process that is contextually and clinically meaningful. At each layer l, we compute the aggregated message $m_{p_i}^{(l)}$ for node p_i as:

$$m_{p_i}^{(l)} = m_{\circlearrowleft p_i}^{(l)} + m_{\rightarrow p_i}^{(l)}, \tag{1}$$

where $m_{\rightarrow p_i}^{(l)}$ and $m_{\circlearrowleft p_i}^{(l)}$ are the contributions from neighbors and self representation, respectively. The neighbor message term $m_{\rightarrow p_i}^{(l)}$ is defined as:

$$m_{\rightarrow p_i}^{(l)} = \sum_{r \in \mathcal{R}_p} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} \mathbf{W}_r^{(l)} h_{p_j}^{(l)}, \tag{2}$$

where $\mathcal{R}_p$ denotes the set of relation types in $\mathcal{G}_p$, $\mathcal{N}_i^r$ is the set of r-type neighbors of node p_i, $W_r^{(l)}$ is a relation-specific transformation matrix, and $c_{i,r}$ is a normalization constant (based on the relation-wise degree). The self-information signal term is defined as:

$$m_{\circlearrowleft p_i}^{(l)} = \mathbf{W}_0^{(l)} h_{p_i}^{(l)}, \tag{3}$$

where $\mathbf{W}_0^{(l)}$ is a learned transformation for the node's self-embedding. The updated node representation $h_{p_i}^{(l+1)}$ is then computed by applying a nonlinear activation function to the aggregated message: $h_{p_i}^{(l+1)} = \sigma\left(m_{p_i}^{(l)}\right)$, where σ is a pointwise nonlinearity such as the sigmoid function. This two-step formulation separates the accumulation of information from relational neighbors and the transformation into a new representation space. This separation improves model interpretability and modularity while preserving the inductive relational biases critical for downstream clinical tasks (as shown in Figure 2).

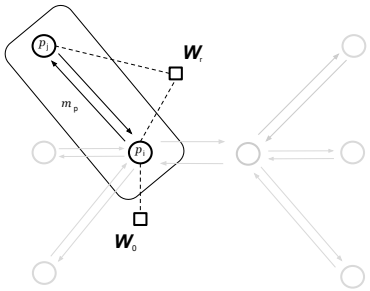


Figure 2: Patient messages between node p_i and node p_j shown as circles and trainable parameters shown as squares.

2.2 ONTOLOGY GRAPH MESSAGE COMPUTATIONS

The ontology graph $\mathcal{G}_q$ encodes structured clinical knowledge in the form of hierarchical concepts and relations. The representations for ontological concepts are learned using hyperbolic neural message passing over an L layer graph neural network with $\kappa_2 = 1$ corresponding to the Poincaré model. In the Poincaré model, embeddings lie in the curved space $\mathbf{B}_c^d = \{x \in \mathbb{R}^d : \|x\| < c\}$ with $c = 1$, well-suited for representing hierarchical structures Liu et al. (2019). A set of directed edges $\mathcal{R}_q$ shapes the flow of information by guiding traversals across the graph's semantic structure, influencing how node features are aggregated and propagated within the ontology. At each layer l, the aggregated message $m_{q_i}^{(l)}$ for node q_i is computed as:

$$m_{q_i}^{(l)} = m_{\circlearrowleft q_i}^{(l)} + m_{\rightarrow q_i}^{(l)}, \tag{4}$$

where $m_{\rightarrow q_i}^{(l)}$ and $m_{\circlearrowleft q_i}^{(l)}$ are the contributions from neighbors and self representation, respectively. The neighbor message term is computed by:

$$m_{\rightarrow q_i}^{(l)} = \sum_{r \in \mathcal{R}_q} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} \exp_{x'}^B\left(\mathbf{W}_r^{(l)} \log_{x'}^B(h_{q_j}^{(l)})\right), \tag{5}$$

where $\mathcal{R}_q$ are relation types, $\mathcal{N}_i^r$ are r-type neighbors, and $\mathbf{W}_r^{(l)}$ is a relation-aware operator, and $c_{i,r}$ is a normalization constant (based on the relation-wise degree). The self-information message is computed by:

$$m_{\circlearrowleft q_i}^{(l)} = \exp_{x'}^B\left(\mathbf{W}_0^{(l)} \log_{x'}^B(h_{q_i}^{(l)})\right), \tag{6}$$

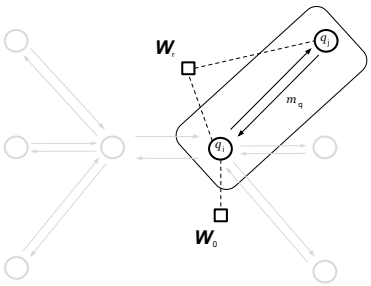


Figure 3: Ontology messages between node q_i and node q_j shown as circles and trainable parameters shown as squares.

where $\mathbf{W}_0^{(l)}$ is a learned transformation for the node's self-embedding. The updated node representation is then given by: $h_{q_i}^{(l+1)} = \sigma\left(m_{q_i}^{(l)}\right)$, where σ is a pointwise nonlinearity such as the $\tanh(h)$ function applied in tangent space. Projection and mapping are done using $\log^B(.)$ and $\exp^B(.)$. Mobius addition and multiplication are used to combine terms. By explicitly separating self-information and neighbor aggregation through hyperbolic transformations, this design captures both local and relational information while preserving hierarchical structure (as shown in Figure 3).

2.3 Alignment Graph Message Computations

Alignment between the encoded representations in the patient graph $\mathcal{G}_p$ and the ontology graph $\mathcal{G}_q$ is achieved through inter-graph message aggregation and transformation. The component is aimed to transfer ontological knowledge over to the clinical domain. For a node p_i in the patient graph, the incoming message aggregation from its ontology neighbors at layer l is computed by:

$$m'^{(l)}_{p_i} = m^{(l)}_{\circlearrowleft p_i} + m'^{(l)}_{\rightarrow p_i}, \tag{7}$$

involving self-information message $m^{(l)}_{\circlearrowleft p_i}$ defined as before and the inter-graph message term given by:

$$m'^{(l)}_{\rightarrow p_i} = \sum_{j \in \mathcal{N}'_i} \frac{1}{c_i} W^{(l)}_s \log^B_{x'}(h^{(l)}_{q_j}), \tag{8}$$

where $W^{(l)}_s$ denotes the alignment weight matrix, $\mathcal{N}'_i$ the set of aligned ontology neighbors, and c_i a normalization constant (e.g., based on degree distribution). Similarly, for a node q_i in the ontology graph, the incoming message from patient neighbors at layer l is computed by:

$$m'^{(l)}_{q_i} = m^{(l)}_{\circlearrowleft q_i} + m'^{(l)}_{\rightarrow q_i}, \tag{9}$$

involving self-information message $m^{(l)}_{\circlearrowleft q_i}$ defined as before and the inter-graph message term given by:

$$m'^{(l)}_{\rightarrow q_i} = \sum_{j \in \mathcal{N}'_i} \frac{1}{c_i} \exp^B_{x'}(W^{(l)}_s h^{(l)}_{p_j}), \tag{10}$$

where $W^{(l)}_s$ denotes the alignment matrix matrix, $\mathcal{N}'_i$ the set of aligned patient neighbors, and c_i a normalization constant (e.g., based on degree distribution). The updated representations are then computed by applying a nonlinearity to the aggregated messages $h^{(l+1)}_{p_i} = \sigma\left(m'^{(l)}_{p_i}\right)$ and $h^{(l+1)}_{q_i} = \sigma\left(m'^{(l)}_{q_i}\right)$, where σ denotes a pointwise nonlinearity such as the sigmoid or tanh function. These cross-graph updates ensure bidirectional alignment by exchanging mutual information between matched nodes in both graphs (as shown in Figure 4).

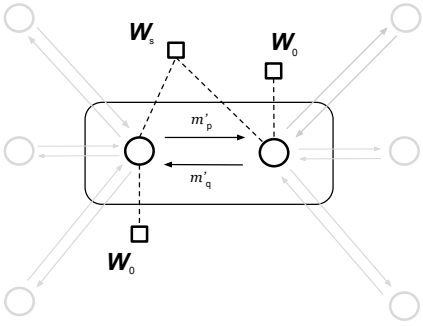


Figure 4: Alignment messages between patient node p_i and ontology node q_i shown as circles and trainable parameters shown as squares.

2.4 Concurrent Predictions

In the remaining part of this section, we introduce an objective function to combine task-specific, generalizable, and ontological features across patient and ontology graphs using the following formulations:

DistMult Decoder: over the patient graph $\mathcal{G}_p$, we employ a DistMult decoder to score triples (s, r, t) using a margin-based ranking loss:

$$\mathcal{L}^p_{\text{link}} = \frac{1}{|\mathcal{T}|} \sum_{(s,r,t)\in\mathcal{T},(s',r,t')\in\mathcal{T}_{\text{neg}}} \max(0, \gamma - f(s,r,t) + f(s',r,t')), \tag{11}$$

where the DistMult scoring function is defined as $f(s,r,t) = \langle \mathbf{h}_s, \mathbf{h}_r, \mathbf{h}_t \rangle$, with $\mathbf{h}_s$ and $\mathbf{h}_t$ denoting node embeddings and $\mathbf{h}_r$ the relation embedding. Here, γ is the margin (set to 1.0), $\mathcal{T}$ is the set of positive triples, and $\mathcal{T}_{\text{neg}}$ denotes corresponding negative triples. We perform a stratified edge split to construct training and validation sets while preserving the distribution of relation types. Positive edges are divided according to the specified train–validation ratio, and an equal number of negative edges are generated by randomly corrupting either the source or destination node of each positive edge. This ensures balanced, relation-aware datasets suitable for link prediction evaluation.

Möbius Decoder: over the ontology graph $\mathcal{G}_q$, we employ a Möbius decoder to score triples (s, r, t) using a margin-based ranking loss:

$$\mathcal{L}^q_{\text{link}} = \frac{1}{|\mathcal{T}|} \sum_{(s,r,t)\in\mathcal{T},(s',r,t')\in\mathcal{T}_{\text{neg}}} \max(0, \gamma - \varphi(s,r,t) + \varphi(s',r,t')),$$

where the Möbius scoring function is defined as $\varphi(s,r,t) = -d_{\mathbf{B}}\big(\mathbf{e}^{(r)}_s, \mathbf{e}^{(r)}_t\big)^2$, with $\mathbf{e}^{(r)}_s = \exp^c_0\big(\mathbf{W}_r \log^c_0(\mathbf{h}_s)\big)$ is the translation in the tangent space at the origin and $\mathbf{e}^{(r)}_t = \mathbf{h}_t \oplus_c \mathbf{h}_r$. Here, $\mathbf{h}_s, \mathbf{h}_t \in \mathbf{B}^d_c$ are node embeddings from the hyperbolic graph neural network, b_r is a relation-specific bias, and $d_{\mathbf{B}}$ denotes hyperbolic distance in

the Poincare ball $\mathbf{B}_c^d$. The margin γ is set to 1.0, $\mathcal{T}$ is the set of positive triples, and $\mathcal{T}_{\text{neg}}$ denotes corresponding negatives. Stratified edge splits are applied to preserve relation-type distributions, with negatives generated by randomly corrupting source or target nodes.

Visit-level prediction: in the patient survival task (i.e. mortality prediction), each visit $v_p^t \in \mathcal{V}_p$ for patient p is represented as a subgraph $\mathcal{G}_p^t = (\mathcal{C}_p^t, \mathcal{E}_p^t)$, where $\mathcal{C}_p^t \subseteq \mathcal{C}_p$ denotes the medical codes (observations, diagnoses, or substances) associated with the visit, and $\mathcal{E}_p^t$ are their induced relations. Message passing constrained within each visit subgraph, ensures that embeddings capture intra-visit structure. After L layers of message passing, we obtain embeddings $\{\mathbf{h}_c \mid c \in \mathcal{C}_p^t\}$ for the codes in visit v_p^t. A visit-level representation is then computed using a permutation-invariant pooling operator: $\mathbf{h}_p^t = f\big(\{\mathbf{h}_c : c \in \mathcal{C}_p^t\}\big)$, where $f(.)$ can be mean pooling, max pooling, or attention-based aggregation. The pooled embedding $\mathbf{h}_p^t$ is mapped to a binary prediction of patient survival after visit t: $\hat{y}_p^t = \sigma(\mathbf{w}^\top \mathbf{h}_p^t + b)$, where $\sigma(\cdot)$ is the logistic sigmoid, $\mathbf{w}$ and b are learnable parameters, and $\hat{y}_p^t \in [0, 1]$ is the predicted probability of survival. The model is trained with binary cross-entropy loss:

$$\mathcal{L}_{\text{visit}} = -\frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(p,t)\in\mathcal{D}_{\text{train}}} \Big(y_p^t \log \hat{y}_p^t + (1 - y_p^t) \log(1 - \hat{y}_p^t) \Big),$$

where $y_p^t \in \{0, 1\}$ is the ground-truth survival label for patient p at visit t. This formulation encourages the model to embed visit-level towards survival outcome prediction.

Concurrent prediction loss: to jointly optimize link and visit-level predictions, we define a *concurrent loss* as the sum of the three individual objectives:

$$\mathcal{L}_{\text{concurrent}} = \mathcal{L}_{\text{link}}^p + \mathcal{L}_{\text{link}}^q + \mathcal{L}_{\text{visit}}. \tag{12}$$

By aggregating these losses, the model is encouraged to learn embeddings that are simultaneously discriminative at each task and relationally consistent across edges. This unified objective allows the encoder to capture complementary information across the different prediction tasks, leading to more robust and generalizable representations.

Algorithm 1 Hyperbolic Neural Message Passing

Input: Patient graph $\mathcal{G}_p$, Ontology graph $\mathcal{G}_q$, alignment pairs $\mathcal{A}$.
Output: Updated messages $m_{p_i}^{(l)}, m_{q_i}^{(l)}$.

for each patient node p_i in parallel **do** ▷ patient updates
  Compute self-message $m_{\circlearrowleft p_i}^{(l)}$ by Eq. (3)
  Compute relational message $m_{\to p_i}^{(l)}$ by Eq. (2)
  Aggregated message: $m_{p_i}^{(l)} \leftarrow m_{\circlearrowleft p_i}^{(l)} + m_{\to p_i}^{(l)}$
end for

for each ontology node q_i in parallel **do** ▷ ontology updates
  Compute self-message $m_{\circlearrowleft q_i}^{(l)}$ by Eq. (6)
  Compute relational message $m_{\to q_i}^{(l)}$ by Eq. (5)
  Aggregated message: $m_{q_i}^{(l)} \leftarrow m_{\circlearrowleft q_i}^{(l)} + m_{\to q_i}^{(l)}$
end for

for each pair $(p_i, q_j) \in \mathcal{A}$ in parallel **do** ▷ alignment updates
  Compute patient cross-message $m_{\to p_i}^{\prime(l)}$ by Eq. (8)
  Aggregated message: $m_{p_i}^{\prime(l)} \leftarrow m_{\circlearrowleft p_i}^{(l)} + m_{\to p_i}^{\prime(l)}$
  Compute ontology cross-message $m_{\to q_j}^{\prime(l)}$ by Eq. (10)
  Aggregated message: $m_{q_j}^{\prime(l)} \leftarrow m_{\circlearrowleft q_j}^{(l)} + m_{\to q_j}^{\prime(l)}$
end for

In order to optimize the concurrent objective, node and edge representations are iteratively updated (see Algorithm 1) in the patient graph $\mathcal{G}_p$ and ontology graph $\mathcal{G}_q$ per layer, leveraging the alignment set $\mathcal{A}$. Nodes aggregate two types of information: a self-message that encodes their own state, and a relational message that captures structural dependencies from connected nodes. For each patient graph node p_i, the self-messages are computed based on equation 3 and relational messages (via equation 2, aggregated as m_{p_i}. Similarly, for ontology graph nodes q_i, it computes and aggregates self (via equation 6) and relational messages (via equation 5) into m_{q_i}. In addition, alignment pairs $(p_i, q_j) \in \mathcal{A}$ enable cross-graph message passing, where patient and ontology nodes exchange cross-messages to refine their embeddings (via equation 8 and equation 10). This joint propagation enriches representations with both intra-graph and inter-graph knowledge, supporting downstream tasks such as outcome prediction, classification, and clinical decision support.

3 EXPERIMENTS

In this section, we conduct experiments under a concurrent prediction framework, where structure-less and hierarchical objectives as well as the visit-level loss function are jointly optimized through the above unified loss function (see relation 12). The experiments are conducted on three sets of real-world EHR datasets to investigate the predictive power of the proposed framework in practical settings. We compare the results against existing methods, including non-Euclidean and EHR specific models, to further understand the different aspects of hybrid, concurrent representation learning. We also include ablation results to highlight advantages from proposed architectural changes in the following subsections.

Datasets: we use three sets of EHR data, namely MIMIC-III Johnson et al. (2016), MIMIC-IV Johnson et al. (2021), and eICU Pollard et al. (2018), focusing on patients hospitalized for Cardiovascular Disease (CVD) disease based on ICD codes. The scope includes medication, hospital procedures, diagnoses, patient demographics, and lab results. MIMIC-III includes 38,597 adult patients and 49,785 admissions, using ICD-9 terminology for diagnosis; MIMIC-IV covers a more recent period with improved data quality, using both ICD-9 and ICD-10

terminologies; and eICU comprises ICU records from multiple U.S. sites (2014–2015), with diagnosis, medication, and procedure codes. Clinical codes (e.g. ICD, NDC, LOINC) are mapped to unique identifiers and aligned with SNOMED CT[1] as the ontology graph of choice, resolving ambiguities with conjunctive relations (and) and membership relations (is a) to enhance contextual knowledge for downstream tasks (see Appendix for details).

Baselines: we consider different baseline models in order to investigate the various aspects of the proposed framework ranging from graphical and EHR specific to manifold learning models, that include:

- Meta-path Models (M): in terms of graphical methods that rely on generating meta-paths for structured data, Metapath2vec Dong et al. (2017) introduces meta-path based random walks and feature learning via skip-gram model; MAGNN Fu et al. (2020) further introduces intra-metapath and inter-metapath aggregation for neighborhood selection; and HSGNN Liu et al. (2020) proposes heterogeneous representation learning via similarity graphs.
- Graph Transformers (G): to capture local and global interactions, Graphormer Ying et al. (2021) encodes graph-structured information using a Transformer architecture; and GraphGPS Rampášek et al. (2022) combines local message passing with global attention.
- EHR Models (K): in terms of models focusing specifically on EHR data, GRAM Choi et al. (2017a) captures ancestor relationships among medical concepts via graph attention; EHR2HG Sun et al. (2022) applies hyper graph convolution for capturing higher order patient-disease relations; and MDP-GRL Guo et al. (2025) applies multi-hop neighbor feature aggregation using patient knowledge graphs for contextual enrichment.
- Hyperbolic (H), Euclidean (E), Dual-space (D): in terms of learning on manifolds, HGCGE Bao et al. (2025) applies graph convolution operations in the hyperbolic space; AttH Chami et al. (2020) combines reflection and symmetric embeddings based on hyperbolic transformations for link prediction over heterogeneous knowledge graph; and GIE Cao et al. (2022) learns knowledge graph representations through interactions between Euclidean, Hyperbolic, and Hyper-spherical spaces.

Training: all models are trained using $\mathcal{L}_{concurrent}$ objective function in relation 12 and the adaptive stochastic gradient optimizer Kingma & Ba (2014) with a learning rate of 0.01 for 10 epochs or using the Riemann Adam Bécigneul & Ganea (2018) to conform to the curvature of the manifold in case of hyperbolic spaces. To ensure statistical robustness, training and evaluation were repeated across 10 independent runs, each with randomly re-initialized model parameters. In each run, a subset of 100 visits is selected for training and evaluation at random based on the medical codes from each visit. All models are trained on 80% of data and tested on 20% of data separately held, with node and edge stratification, and negative samples generated uniformly at random. The RGCN was initialized with 2 layers and hidden dimensionality of 64. MAGNN was trained using metapath-based aggregation with 64 hidden units, 32-dimensional attention vectors, and 2 attention heads. The GPSTransformer (dim = 32, PE = 4, RW-PE = 10) was configured with 1 attention head and 1 layer and GraphTransformer was configured with 2 layers, 4 attention heads, and 8 spectral features. Metapath2Vec, HSGNN and EHR2HG models used hidden dimensions of 64, 32 and 32, respectively. All hyperbolic models are set to have a hidden dimension of 16. For concurrent training, our proposed model is set with $\kappa_1 = 0$ and $\kappa_2 = 1$, while HGCGE and AttH are trained with $c = 1$ as curvature parameter.

3.1 Performance using Concurrent Learning

We evaluate and report the performance of the proposed hyperbolic neural message passing model over EHR data, named HNMP-EHR, using precision, recall, and F1 score on multiple tasks, including mortality classification, diagnosis outcome prediction, lab even detection, and medication recommendation. The results are summarized in Table 1. Overall it can be observed that due to the specific nature of EHR data wherein the tree-like visit records are highly structure-less whereas the ontological information are highly structured, the proposed model can achieve the most gain in performance by learning over independent embedding spaces and subsequently aligning the representations thorough cross message passing. Methods without an explicit mechanism to do so, seemingly fail to capture structural disparities present in EHR datasets. To this effect, our proposed model performs consistently across all selected datasets and among the different model categories, further elaborated in what follows.

Among general purpose and EHR specific models, high performance is achieved on single tasks however by different methods (details can be found in the Appendix). In terms of individual performance over MIMIC III and MIMIC IV visits, EHR2HG for Diagnosis prediction and Lab test detection, as well as, Metapath2Vec for medication recommendation and mortality prediction stand out with F1 scores over 0.90 points that is acceptable in practical settings. Over eICU dataset, meta-path based models including MAGNN and Metapath2Vec similarly achieve near perfect performance on a given subset of tasks, however not at once. In this category, models that

[1] https://www.snomed.org/

Table 1: Concurrent performance of HNMP-EHR (highlighted) in comparison with baseline models across MIMIC III, MIMIC IV, and eICU datasets in terms of Precision, Recall, and F1 score. In each type category, the best and second best F1 scores are boldfaced and underscored respectively. In each case, the mean value and standard deviation averaged over 10 independent runs are reported. Type categories are: Meta-path Models (M), Graph Transformers (G), EHR Models (K), Hyperbolic (H), Euclidean (E), and Dual-space (D).

		MIMIC III			MIMIC IV			eICU		
Method	**Type**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
GraphGPS	G	0.87±0.25	0.58±0.05	0.65±0.14	0.76±0.31	0.51±0.14	0.58±0.22	0.81±0.38	0.47±0.13	0.55±0.26
GRAM	K	0.79±0.35	0.72±0.33	0.73±0.33	0.84±0.32	0.53±0.38	0.56±0.36	0.80±0.39	0.41±0.15	0.51±0.25
MDP-GRL	K	0.79±0.42	0.53±0.32	0.61±0.34	0.79±0.42	0.57±0.29	0.64±0.34	0.83±0.33	0.57±0.25	0.66±0.29
HSGNN	M	0.79±0.42	0.65±0.35	0.69±0.36	0.82±0.36	0.73±0.33	0.76±0.34	0.68±0.28	0.51±0.44	0.53±0.41
Graphormer	G	0.79±0.41	0.69±0.35	0.70±0.37	0.81±0.38	0.67±0.33	0.71±0.34	0.81±0.38	0.50±0.28	0.58±0.31
MAGNN	M	0.85±0.29	0.73±0.31	**0.77±0.28**	0.82±0.34	0.76±0.30	0.78±0.32	0.85±0.16	0.73±0.39	0.74±0.36
EHR2HG	K	0.72±0.48	0.73±0.46	0.72±0.47	0.72±0.48	0.73±0.46	0.72±0.48	-	-	-
Meta2Vec	M	0.93± 0.13	0.68±0.37	0.72±0.31	0.95±0.09	0.80±0.22	**0.85±0.17**	0.99±0.01	0.85±0.23	**0.90±0.15**
RGCN	E	0.98±0.04	0.87±0.02	**0.91±0.02**	0.98±0.03	0.79±0.23	**0.85±0.18**	0.99±0.01	0.89±0.15	**0.93±0.10**
HGCGE	H	0.29±0.23	0.34±0.11	0.26±0.19	0.31±0.13	0.35±0.13	0.29±0.14	0.33±0.28	0.40±0.15	0.33±0.21
AttH	H	0.79±0.42	0.72±0.36	0.75±0.38	0.77±0.44	0.64±0.32	0.69±0.39	0.79±0.42	0.67±0.34	0.72±0.37
GIE	D	0.79±0.42	0.73±0.36	0.76±0.38	0.78±0.42	0.67±0.35	0.71±0.37	0.79±0.42	0.73±0.36	0.75±0.38
HNMP-EHR	D	1.00±0.00	0.94±0.05	**0.96±0.03**	0.99±0.01	0.94±0.09	**0.96±0.05**	1.00±0.00	0.91±0.12	**0.95±0.10**

overall rely on global information transfer (including MAGNN and Meta2Vec with concurrent F1 scores of 0.77, 0.86 and 0.90 on MIMIC III, MIMIC IV, and eICU respectively), stand out among general purpose embedding models.

Among models that capture global and local dependencies using the transformer architecture, the concurrent performance is on average the lowest as compared to other categories. This performance drop is in part due to the diversity of tasks that require multiple levels of structural understanding, and in part due to the mixed nature of EHR data. While these models excel at intra-visit inference, the inter-visit information critical for mortality prediction for example is lost in the process, causing a significant performance drop (details can be found in the Appendix). For instance, the GTransformer model achieves F1 scores of 0.98 and 0.99 for diagnosis prediction and medication recommendation, but its mortality prediction rate drops to 0.18 points on MIMIC III dataset. Similar performance degradation is observed on other datasets and by GraphGPS. Without additional architectural modifications (e.g. recurrent prediction heads), transformer-based models therefore are unable to disambiguate mixed type interactions. Furthermore, when it comes to intra-visit interactions alone and the additional hierarchical information provided through the ontological embeddings, the attention mechanism again fails to deduce such connections.

Among models that utilize manifold learning through message passing specifically, the AttH outperforms the same type of model on Diagnosis prediction and ties with GIE on Lab prediction which achieves the highest performance on Medication recommendation. Although these models learn structurally sound representations including hierarchical information, their performance degrades overall, specially on tasks such as mortality prediction where graph-level representations and chain-like data structures play an important role. The RGCN which utilizes message passing over the Euclidean space is able to achieve over 0.90 points across diverse tasks in two of datasets, namely MIMIC III and eICU. Its performance overall is however limited due to the lack of an explicit mechanism to handle hierarchical information, as evident in the results. On the other hand, HNMP-EHR (ours) which uses a separate embedding space to capture mixed types of interactions, overcomes this barrier for concurrent prediction. This is believed to be the due to the ability of the model to learn and exchange knowledge from both structure-less (i.e. EHR) and ontological (i.g. SNOMED CT) information, efficiently.

3.2 Effect of Dual-Space Alignment

In heterogeneous representation learning, alignment plays a central role in bridging different sources of information. While the patient graph contains limited inherent structure, the ontology graph provides a rich hierarchy of relationships defined by subclass structures and transitive closures. By carefully aligning these two graphs, structured knowledge can be transferred into patient representations, enabling more meaningful embeddings and improving task performance. This controlled transfer ensures that knowledge is shared in order to reinforce the overall representational capacity across applications. To assess this effect on performance, we showcase its impact by adjusting the hyper-parameters κ_1 and κ_2 to control the amount of hierarchical representation shared in each respective space. According to the properties of hyperbolic geometry, by varying these parameters that are related to the curvature of the underlying manifold, different degrees of hierarchical information can be learned.

In particular, increasing the κ parameter in each space from zero towards one, will increase the amount of hierarchical information encoding. In our model, when the hyper parameters are set as $\kappa_1 = \kappa_2 = 0$, the embeddings lie on the Euclidean plane (E) corresponding to the case in which the representations are learned using message passing with a relational graph neural network (with a specific edge type that connects nodes in the alignment graph $\mathcal{G}_\mathcal{A}$). Conversely, setting the hyper parameter as $\kappa_1 = \kappa_2 = 1$ causes the embeddings to lie on a curved manifold, corresponding to the case for message passing with a hyperbolic graph neural network and a hyper-sphere of curvature one (P). In the third scenario, with the hyper parameter $\kappa_1 = 0$ over patient graph $\mathcal{G}_p$ and the hyper parameter $\kappa_2 = 1$ over ontology graph $\mathcal{G}_q$, the EHR representations are learned in the Euclidean space, while the ontology representations are learned in the Hyperbolic space (PE). Since the patient graph is relatively structure-less, the choice of $\kappa_1 = 0$ is suitable using the Euclidean metric as the scoring function, which is able to capture tree-like structures related to patient visits. Conversely, the choice of $\kappa_2 = 1$ is better suited to the type of representations over the ontology graph nodes which are more hierarchically oriented. In combination, as shown in Figure 5 (top), the hybrid representations (PE) are observed to improve performance over settings where a mono curve manifold is learned (E or P). More specifically, the curvature κ_2 impacts the amount of hierarchical information that is learned from $\mathcal{G}_q$ and (bidirectionally) transferred to $\mathcal{G}_p$ via messages m'_p (and m'_q). With a fixed κ_1, as the value of κ_2 is increased from zero towards one, the amount of learned ontological information also increases as shown in Figure 5 (bottom). At one extreme (i.e. $\kappa_2 = 0$), the amount of hierarchical information transferred between dual representations is minimal, roughly corresponding to the case of flat manifold learning. However, in the other extreme (i.e. $\kappa_2 = 1$), the underlying structures of disparate graphs are found to enhance concurrent performance through mutual messages.

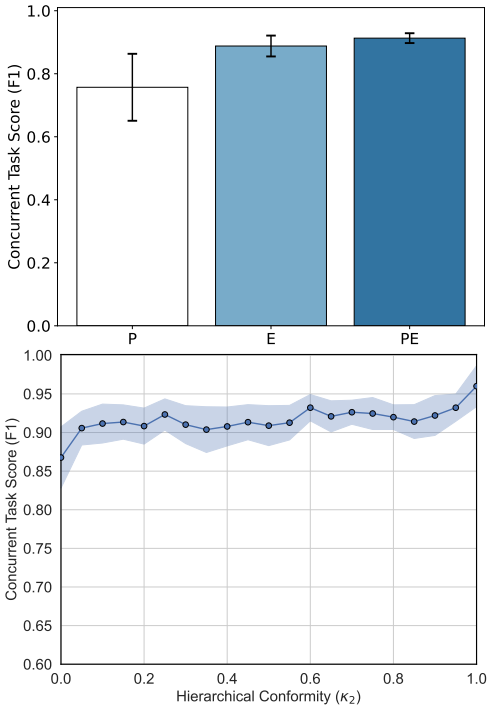


Figure 5: Poincaré (P), Euclidean (E), and dual model (PE) concurrent task performances (top) and the effect of hierarchical conformity (bottom).

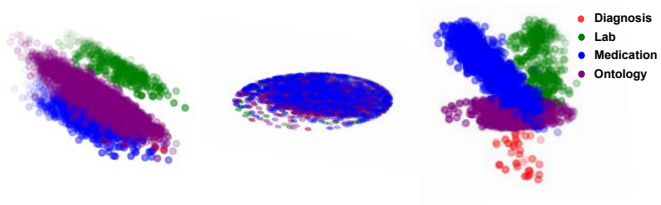
4 3D Visualization of Ontology-aligned Representations

In the last set of experiments we inspect the properties of hyperbolic neural message passing by visualizing in the three dimensional space, the learned representations on a subset of $\mathcal{G}_p$ with diagnosis, observations, and medication nodes, as well as their corresponding alignments from the ontology graph $\mathcal{G}_q$. The model in PE setting ($\kappa_1 = 0$ and $\kappa_1 = 1$) is retrained with hidden dimension $d = 3$, and applied to the subset of codes as shown in Figure 6. We also generate embeddings using the E parameter setting ($\kappa_1 = 0$ and $\kappa_1 = 0$) and P parameter setting ($\kappa_1 = 1$ and $\kappa_1 = 1$) on the same subset of nodes. It is evident that the Euclidean (E) manifold demonstrates adequate task-specific separability of nodes (achieving a F1 score of 0.90), while the Poincare (P) model's embeddings lie onto the single hyperbolic disk $\mathbf{B}_1^3$ (corresponding to a decrease in concurrent performance to 0.74 points). On the other hand, better separability can be seen using the PE setting whereby the embedding planes of diagnosis, lab, and medication embeddings (shown in red, green, and blue) and the Poincare ball hierarchies (shown in purple) experience better separability (which explains improved performance in F1 score to 0.96). As evident, the mutual knowledge that is transferred through alignment is crucial for improving the quality of EHR representations. Whereas the patient graph is relatively structureless, the ontology graph encapsulates rich hierarchical knowledge encoded by transitive closures and membership relationships. The transfer of information between spaces as shown therefore is a major component in improving the capacity of the model to extract and apply structured knowledge across multiple tasks.

5 Related Work

A summary of the state-of-the-art methods for multi-purpose applications based on heterogenuous graph learning representations are provided in this section. In Fouladvand et al. (2023), a novel graph neural network (GNN)-based framework is proposed to predict medical **procedures** using electronic health records. By training AI models on large-scale EHR data, their approach provides personalized procedure recommendations, enhancing the efficiency of clinician time, reducing diagnostic delays, and potentially eliminating the need for in-person consultations. For **medication** recommendations, the approach in Shang et al. (2019) models patient records as queries for personalized combinations and Yang et al. (2021) employs recurrent residual networks to

κ_1	κ_2	Precision	Recall	F1 Score
0	0	0.98	0.86	0.90
1	1	0.78	0.72	0.74
0	1	**0.99**	**0.95**	**0.96**



(a) $\kappa_1 = 0, \kappa_2 = 0$ (b) $\kappa_1 = 1, \kappa_2 = 1$ (c) $\kappa_1 = 0, \kappa_2 = 1$

Figure 6: Concurrent prediciton performance (left) and 3D embeddings visualization: Euclidean (E) space with $\kappa_1 = 0, \kappa_2 = 0$, Poincaré (P) space with $\kappa_1 = 1, \kappa_2 = 1$, and dual (PE) space with $\kappa_1 = 0, \kappa_2 = 1$ (right).

predict medication changes, unlike traditional methods requiring full patient history. For the **diagnosis** task, the proposed model in Sun et al. (2021) constructs medical concept and patient record graphs from EHR and external knowledge bases, using a neural graph encoder to generate embeddings for patients, diseases, and symptoms, enabling inductive inference for new patients. The method of Wu et al. (2021) uses a hierarchical framework with biased-random-walk for medical and proximity-preserving embeddings. **Mortality** prediction presents another promising data-driven application, which has recently been the subject of research as a stand-alone task in Jiang et al. (2024). In this work, we propose a unifying framework that can supports the aforementioned applications in a multi-task setting (i.e. concurrent predictions).

Several general-purpose methods have been proposed to address key challenges in **heterogeneous graph representation learning** from a technical perspective. The study in Fu et al. (2020) enhances heterogeneous graph embedding by introducing a threefold approach: transforming node content to include feature-level semantics, aggregating information within metapaths to model structural dependencies among intermediate nodes, and combining multiple metapaths to capture diverse relational patterns. The issue of insufficient relational modeling in heterogeneous graphs is further addressed in Yu et al. (2023). This approach employs cross-relation message passing to enable fine-grained semantic and structural interaction propagation. A number of methods have been proposed that further focus on capturing structural information using **Hyperbolic Graph Neural Networks** (HGNN). The Poincare model in particularly has been shown to learn hierarchical representations well Chami et al. (2020) and applied over structured data including multi property and knowledge graphs Bao et al. (2025). Dual geometric space embedding for knowledge graphs is proposed in Iyer et al. (2022) to jointly represent ontology and specific instances in a unified framework. The approach in Cao et al. (2022) further extends this concept to multi space interactions. In this work, we utilize a geometric learning approach to capture ontological information combined with inter-graph message passing for cross-space interactions with EHR representations.

6 Conclusions

We propose HNMP-EHR[2], a multi-task prediction framework for real-world EHR data, leveraging a hyperbolic neural message computation strategy that learns and transfers structure-dependent knowledge from different embedding spaces. Our approach improves performance by aligning patient data with ontological knowledge via cross-graph message passing. Experiments on real-world datasets showcase that robust representations can be learned for downstream clinical tasks concurrently. For concurrent prediction over real-world datasets, the proposed method achieves consistent scores across diverse tasks, with an average scores of 0.99, 0.95, and 0.96 in precision, recall and f1 score, outperforming mono-geometric and general purpose models including RGCN (with average scores of 0.98, 0.86, and 0.90) and Metapath2Vec (with average scores of 0.95, 0.78, and 0.82). These results highlight HNMP-EHR's ability to capture complex relationships in EHR data where the implicit hierarchy information plays a crucial role.

[2]Our code is available at: `https://github.com/s-hm-a/hnmp_ehr`

7 Ethics Statement

We have read the ICLR Code of Ethics and ensured this paper follows it. Our work does not involve dataset releases; all benchmark datasets are publicly available. We anticipate that our work yields a positive societal impact by advancing primary-care research through improved analysis of publicly available datasets.

References

Liming Bao, Yan Wang, Xiaoyu Song, and Tao Sun. Hgcge: hyperbolic graph convolutional networks-based knowledge graph embedding for link prediction. *Knowledge and Information Systems*, 67(1):661–687, 2025.

Gary Bécigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. *arXiv preprint arXiv:1810.00760*, 2018.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. Geometry interaction knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 5521–5529, 2022.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545*, 2020.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. Gram: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 787–795, 2017a.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. GRAM: Graph-based Attention Model for Healthcare Representation Learning. *KDD : proceedings. International Conference on Knowledge Discovery & Data Mining*, 2017:787–795, August 2017b. ISSN 2154-817X. doi: 10.1145/3097983.3098126. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7954122/.

Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, and Andrew Dai. Learning the Graphical Structure of Electronic Health Records with Graph Convolutional Transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):606–613, April 2020. ISSN 2374-3468. doi: 10.1609/aaai.v34i01.5400. URL https://ojs.aaai.org/index.php/AAAI/article/view/5400. Number: 01.

Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 135–144, 2017.

Lucas Ferraz, Pedro Giesteira Cotovio, and Catia Pesquita. Can Language Models Align Biomedical Ontologies?: Evaluating Retrieval-Augmented Prompt Strategies in Bio-ML. *8th Workshop on Semantic Web solutions for large-scale biomedical data analytics*, June 2025. URL https://ceur-ws.org/Vol-4001/paper5.pdf.

Sajjad Fouladvand, Federico Reyes Gomez, Hamed Nilforoshan, Matthew Schwede, Morteza Noshad, Olivia Jee, Jiaxuan You, Rok Sosic, Jure Leskovec, and Jonathan Chen. Graph-based clinical recommender: Predicting specialists procedure orders using graph representation learning. *Journal of Biomedical Informatics*, 143:104407, July 2023. ISSN 1532-0464. doi: 10.1016/j.jbi.2023.104407. URL https://www.sciencedirect.com/science/article/pii/S1532046423001284.

Bryony Dean Franklin, Charles Vincent, Mike Schachter, and Nick Barber. The Incidence of Prescribing Errors in Hospital Inpatients. *Drug Safety*, 28(10):891–900, October 2005. ISSN 1179-1942. doi: 10.2165/00002018-200528100-00005. URL https://doi.org/10.2165/00002018-200528100-00005.

Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding, March 2020. URL http://arxiv.org/abs/2002.01680. arXiv:2002.01680 [cs].

Yongan Guo, Yeqi Huang, Yuao Wang, Yun Liu, Shenqi Jing, Tao Shan, Yuan Miao, and Bo Li. Mdp-grl: Multi-disease prediction by graph-enabled representation learning. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2025. doi: 10.1109/JBHI.2025.3584916.

Junheng Hao, Chuan Lei, Vasilis Efthymiou, Abdul Quamar, Fatma Özcan, Yizhou Sun, and Wei Wang. Medto: Medical data to ontology matching using hybrid graph neural networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pp. 2946–2954, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467138. URL https://doi.org/10.1145/3447548.3467138.

Roshni G Iyer, Yunsheng Bai, Wei Wang, and Yizhou Sun. Dual-geometric space embedding model for two-view knowledge graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 676–686, 2022.

Pengcheng Jiang, Cao Xiao, Minhao Jiang, Parminder Bhatia, Taha Kass-Hout, Jimeng Sun, and Jiawei Han. Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. *arXiv preprint arXiv:2410.04585*, 2024.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Leo Anthony Celi, Roger Mark, and S Horng IV. Mimic-iv-ed. *PhysioNet*, 2021.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1):160035, May 2016. ISSN 2052-4463. doi: 10.1038/sdata.2016.35. URL https://www.nature.com/articles/sdata201635. Publisher: Nature Publishing Group.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. *Advances in neural information processing systems*, 32, 2019.

Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and Philip S Yu. Heterogeneous similarity graph neural network on electronic health records. In *2020 IEEE international conference on big data (big data)*, pp. 1196–1205. IEEE, 2020.

Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 743–752, Torino Italy, October 2018. ACM. doi: 10.1145/3269206.3271701. URL https://dl.acm.org/doi/10.1145/3269206.3271701.

Chengsheng Mao, Liang Yao, and Yuan Luo. MedGCN: Medication recommendation and lab test imputation via graph convolutional networks. *Journal of Biomedical Informatics*, 127:104000, March 2022. ISSN 1532-0464. doi: 10.1016/j.jbi.2022.104000. URL https://www.sciencedirect.com/science/article/pii/S1532046422000168.

Shervin Mehryar. Resolution-alignment-completion of tabular electronic health records via meta-path generative sampling. In *Proceedings of the 4th Table Representation Learning Workshop*, pp. 200–207, 2025.

Rajat Mishra and S. Shridevi. Knowledge graph driven medicine recommendation system using graph neural networks on longitudinal medical records. *Scientific Reports*, 14(1):25449, October 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-75784-5. URL https://www.nature.com/articles/s41598-024-75784-5. Publisher: Nature Publishing Group.

Sameh K Mohamed, Aayah Nounu, and Vít Nováček. Biological applications of knowledge graph embedding models. *Briefings in Bioinformatics*, 22(2):1679–1693, March 2021. ISSN 1477-4054. doi: 10.1093/bib/bbaa012. URL https://doi.org/10.1093/bib/bbaa012.

Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific data*, 5(1):1–13, 2018.

Raphael Poulain and Rahmatollah Beheshti. Graph transformers on ehrs: Better representation improves downstream performance. In *The Twelfth International Conference on Learning Representations*, 2024.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515, 2022.

Sarah Ross, Christine Bond, Helen Rothnie, Sian Thomas, and Mary Joan Macleod. What is the scale of prescribing errors committed by junior doctors? A systematic review. *British Journal of Clinical Pharmacology*, 67(6): 629–640, June 2009. ISSN 1365-2125. doi: 10.1111/j.1365-2125.2008.03330.x.

Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1126–1133, July 2019. ISSN 2374-3468. doi: 10.1609/aaai.v33i01.33011126. URL https://ojs.aaai.org/index.php/AAAI/article/view/3905. Number: 01.

Zhenchao Sun, Hongzhi Yin, Hongxu Chen, Tong Chen, Lizhen Cui, and Fan Yang. Disease Prediction via Graph Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, 25(3):818–826, March 2021. ISSN 2168-2208. doi: 10.1109/JBHI.2020.3004143. URL https://ieeexplore.ieee.org/document/9122573.

Ziyou Sun, Xiaohui Yang, Zhiquan Feng, Tao Xu, Xue Fan, and Jinglan Tian. Ehr2hg: modeling of ehrs data based on hypergraphs for disease prediction. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1730–1733. IEEE, 2022.

Giampaolo P Velo and Pietro Minuz. Medication errors: prescribing faults and prescription errors. *British Journal of Clinical Pharmacology*, 67(6):624–628, June 2009. ISSN 0306-5251. doi: 10.1111/j.1365-2125.2009.03425.x. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2723200/.

Tong Wu, Yunlong Wang, Yue Wang, Emily Zhao, and Yilian Yuan. Leveraging graph-based hierarchical medical entity embedding for healthcare applications. *Scientific Reports*, 11(1):5858, March 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-85255-w. URL https://www.nature.com/articles/s41598-021-85255-w. Publisher: Nature Publishing Group.

Chaoqi Yang, Cao Xiao, Lucas Glass, and Jimeng Sun. Change Matters: Medication Change Prediction with Recurrent Residual Networks. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 3728–3734, Montreal, Canada, August 2021. International Joint Conferences on Artificial Intelligence Organization. doi: 10.24963/ijcai.2021/513. URL https://www.ijcai.org/proceedings/2021/513.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

Le Yu, Leilei Sun, Bowen Du, Chuanren Liu, Weifeng Lv, and Hui Xiong. Heterogeneous graph representation learning with relation awareness. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5935–5947, 2023. doi: 10.1109/TKDE.2022.3160208.

Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph Neural Networks and Their Current Applications in Bioinformatics. *Frontiers in Genetics*, 12, July 2021. ISSN 1664-8021. doi: 10.3389/fgene.2021.690049. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.690049/full. Publisher: Frontiers.

A DATASETS

We use three publicly available repositories containing tabular data for patients to ultimately generate graphs for training and evaluation purposes. In this work, we limit the scope to records from patients that are hospitalized for Cardiovascular Disease (CVD). The relevant data are separated by admissions encoded by ICD-9 code range 410-430, such as 428.22 (Chronic systolic heart failure), 428.23 (Acute on chronic systolic heart failure), 428.32 (Chronic diastolic heart failure), 428.33 (Acute on chronic diastolic heart failure), 428.42 (Chronic combined systolic and diastolic heart failure), and 428.43 (Acute on chronic combined systolic and diastolic heart failure). These codes categorize various forms and severities of heart failure based on the systolic and diastolic dysfunction of the heart. In ICD-10, these codes are largely replaced by categories under I50 (Heart Failure). To generate this subset, we identify and store the admissions for those patients that have at least one of the above ICD codes associated with them and exclude items outside the above scope for our final set of patients. The tabular data used in this work are selected and organized around four core themes, namely Diagnosis, Procedures, Prescriptions, and Lab Events. Although there are cases where extra information such as transfers, provider source, and notes exist, for the purposes of tabular processing related to our use case we organize the data under aforementioned core concepts similar to Poulain & Beheshti (2024); Choi et al. (2020).

MIMIC III - the first version of the repository that we use contains data associated with 53,423 distinct hospital admissions for adult patients (aged 16 years or above) admitted to critical care units between 2001 and 2012. The data covers 38,597 distinct adult patients and 49,785 hospital admissions. The median age of adult patients is 65.8 years, 55.9% patients are male, and in-hospital mortality is 11.5%. In addition, it contains data for 7870 neonates admitted between 2001 and 2008. The median length of an ICU stay is 2.1 days and the median length of a hospital stay is 6.9 days. A mean of 4579 charted observations ('chartevents') and 380 laboratory measurements ('labevents') are available for each hospital admission Johnson et al. (2016).

MIMIC IV - the second version of the repository that we use improves upon the previous version by covering a more recent time period (2008–2019) and offering a modular structure, dividing data into core, hospital, and ICU components for easier access. It expands beyond ICU data to include hospital-wide information, increasing the dataset's scope and diversity. MIMIC-IV also enhances data quality and uses both ICD-9 and ICD-10 coding reflecting disparities in coding versions. These updates make MIMIC-IV a more comprehensive and flexible resource for clinical research compared to MIMIC-III. Among other reasons, we are interested in understanding the effect of version-based variations by including the second repository.

eICU - is a large, deidentified, multi-center dataset containing high-granularity admission data across U.S. hospitals in 2014–2015. It includes vital signs, care plans, severity scores, diagnoses, and treatments, enabling research in critical care, machine learning, and decision support. Clinical information is standardized through ICD-9-CM coding for diagnoses, LOINC identifiers for lab tests, and RxNorm terminology for medications, facilitating interoperability and reproducible research.

In order to generate patient graphs form input relational data, column attributes (e.g. "Diagnosis Code") are mapped to predicates (e.g. "has code"). A patient's record can specifically include coded entities from one or multiple coding systems - ICD for Diagnosis and Procedures, LOINC for Lab Results, and NDC (National Drug Code) for drugs/prescriptions. All clinical codes are transformed and consolidated with unique identifiers to ensure semantic equivalence across sources, as summarized in Table 2. In the case of eICU, the dataset includes 41,026 CVD visits containing 399 diagnosis codes, 191 drug codes, and 188 lab codes. MIMIC-III has 31,839 visits including 149 diagnosis, 969 drug, and 327 lab codes, while MIMIC-IV, the largest, has 147,995 visits, including 386 diagnosis, 861 drug, and 324 lab codes.

In order to disambiguate and contextualize knowledge with unified representations, domain specific alignment of codified data according to clinical and biomedical ontologies is adopted, namely with Systematized Nomenclature of Medicine Clinical Terms[3] (SNOMED CT), that are best suited for medical applications needing high quality and personalized context. SNOMED CT is a comprehensive, multilingual healthcare terminology used to encode clinical information in a consistent and structured way. It enables the representation of clinical concepts such as diagnoses, procedures, and symptoms. SNOMED CT defines clinical terms through structured relations between concepts, enabling detailed descriptions such as 'Pain in the left arm' defined with 'finding site' connected to anatomical site 'left arm'. SNOMED CT features allow users to compose new clinical expressions from existing terms.

[3] https://www.snomed.org/

Table 2: Statistics for each graph type (patient $\mathcal{G}_p$ and ontology $\mathcal{G}_q$) and MIMIC III, MIMIC IV, eICU resources, in terms of the number of medical codes, visits, and extended alignment graphs.

		MIMIC III		**MIMIC IV**		**eICU**	
		$\mathcal{G}_p$	$\mathcal{G}_q$	$\mathcal{G}_p$	$\mathcal{G}_q$	$\mathcal{G}_p$	$\mathcal{G}_q$
Visit Number	$\lvert V_p \rvert$	31,839	-	147,995	-	41,026	-
Diagnosis Codes	$\lvert \mathcal{D}_p \rvert$	149	570	386	2856	399	1580
Drug Codes	$\lvert \mathcal{M}_p \rvert$	969	255	861	840	191	761
Lab Codes	$\lvert \mathcal{O}_p \rvert$	327	458	324	269	188	495
Alignment Graph	$\lvert \mathcal{G}_a \rvert$	–	1360	–	12380	-	13999

Mapping diagnosis and procedure codes to SNOMED CT is done based on label information. Similarly, the National Drug Code (NDC)[4] codes and the Logical Observation Identifiers Names and Codes (LOINC)[5] codes are mapped to relevant concepts from prescriptions and observations in SNOMED CT. For any code with multiple ambiguous matches, each match is added using a conjunctive relation 'and' along with their respective membership graphs through 'is a' relation. The alignment graph will include 1,580 diagnosis, 761 drug, and 495 lab codes in the case of eICU dataset, and the total number of entries 13,999 (including subclass relationships). In the case of MIMIC III, 570 diagnosis, 255 drug, and 458 lab codes are aligned and included for a total alignment graph size of 1,360. In the case of MIMIC IV 2,856 diagnosis, 840 drug, and 269 lab codes are aligned and included for a total of 12,380 alignment entries. The combination of membership and multi-code extensions using SNOMED CT is believed to improve the semantic richness and hierarchical structure of the patient graph, specifically for multi-purpose prediction tasks.

In Figure 7, an example visit graph for a patient with a single diagnosis, two medications, and two lab tests, aligned with a total of eight ontology entries is shown. It should be noted that each patient can and often will contain multiple diagnosis, lab test, and medication codes. Although these codes are identifiable within each visit, the causal and specific relationships between them (e.g. diagnosis leading to medication or test) are not present in the original datasets. One of the key justifications for using ontology alignment in this case is to utilize the hierarchical information to enhance EHR representations. The specific characteristics of visits, in terms of the number of coded entities and the size of subsequent alignment graph, may vary. Further details in terms of the number of unique codes per visit and other dataset statistics are shown in Figure 8.

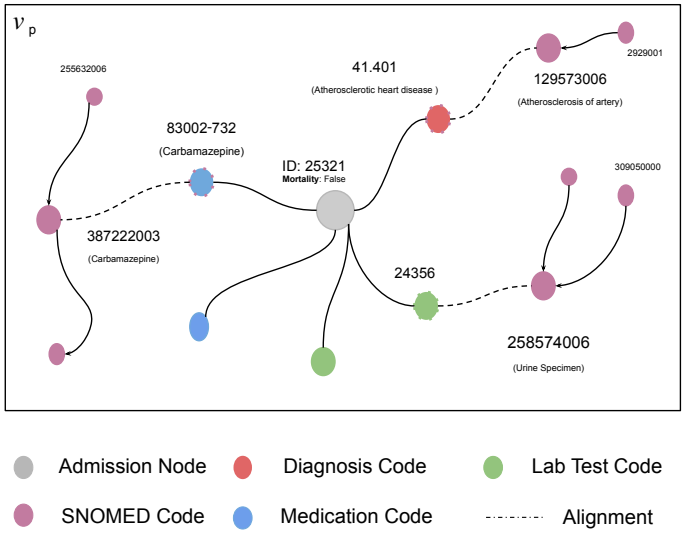
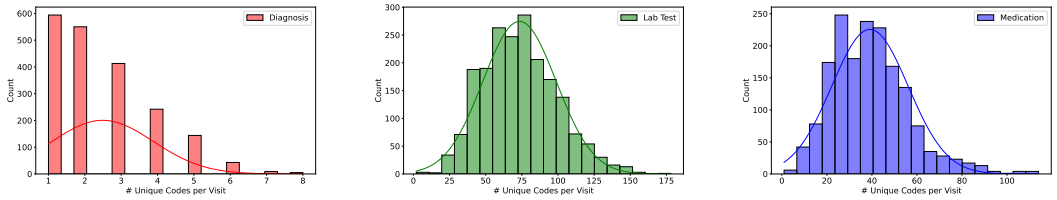
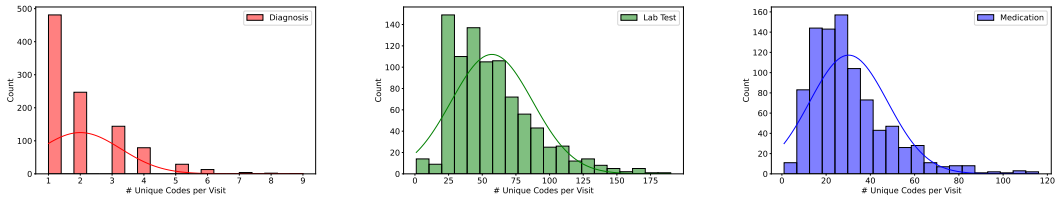


Figure 7: Visit sample v_p for a patient p with diagnosis, medication, and lab tests with three alignment entries to the ontology entities which include membership information using the 'is a' relation, indicated with arrows. The visit node is represented with an ID and a binary mortality label.

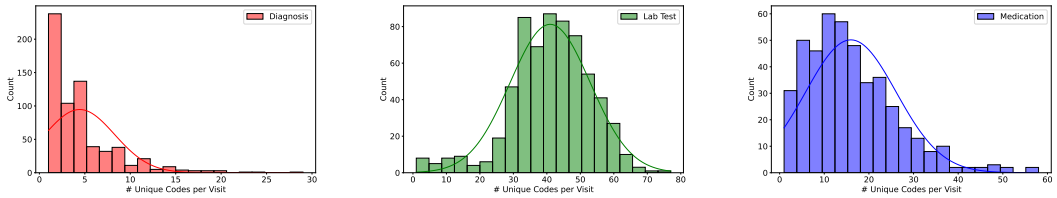
[4] https://dps.fda.gov/ndc
[5] http://loinc.org/



(a) MIMIC III



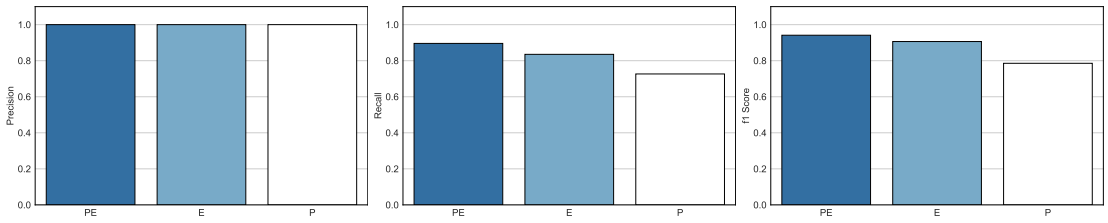
(b) MIMIC IV



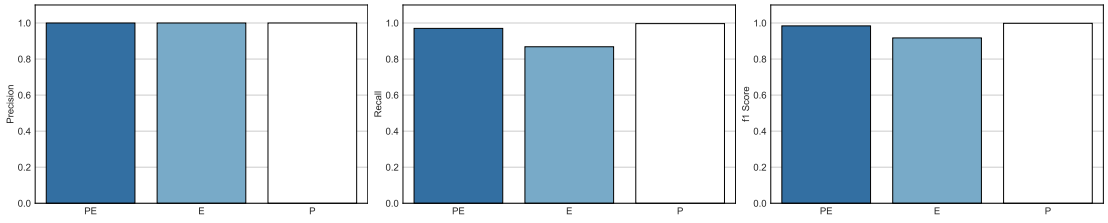
(c) eICU

Figure 8: Distribution of Unique Medical Codes per Patient Visit v_p taken over MIMIC III, MIMIV IV, and eICU Datasets used in order to generate patient graph nodes.

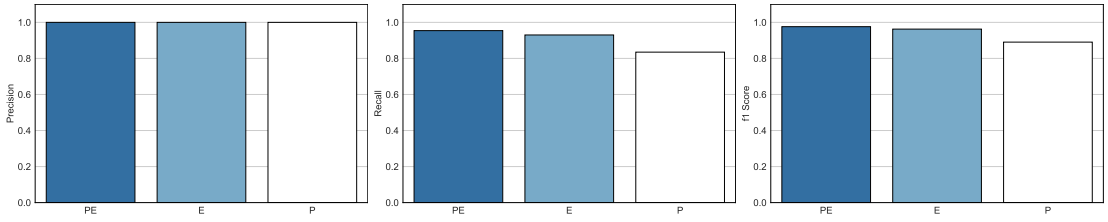
B Performance of Mono- and Dual-Space Hyperbolic Message Passing



(a) MIMIC-III



(b) MIMIC-IV



(c) eICU

Figure 9: Dual model (PE), Euclidean (E), and Poincaré (P) concurrent task performances in terms of Precision (left), Recall (middle), and F1 score (right) on MIMIC III (top), MIMIC IV (middle), and eICU (bottom) datasets.

C SINGLE TASK PREDICTION PERFORMANCE

The mortality task across MIMIC III, MIMIC IV, and eICU datasets, as presented in Table 3, demonstrates the single-task performance of HNMP-EHR on visit-based survavial prediction, achieving F1 scores in the range 0.98-0.99 that is deemed acceptable for practical scenarios. RGCN follows closely, with strong results (0.95–0.99), particularly in eICU. Metapath2Vec performs well in eICU (0.90–0.97) but deteriorates on MIMIC datasets which a different structure. Other methods relying on meta-path and similarity graph approximations, such as MAGNN and HSGNN, show lower than acceptable scores for practical settings. Hyperbolic and dual-space models also tend to under-perform, mortality prediction based on visit data does not heavily depend on hierarchical information deemed appropriate for such methods.

Table 3: Single-task performance for Mortality prediction, on MIMIC III, MIMIV IV, and eICU datasets in terms of Precision, Recall, and F1 score.

	MIMIC III			**MIMIC IV**			**eICU**		
Method	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Metapath2Vec	0.80	0.81	0.81	0.81	0.64	0.67	0.97	0.90	0.92
MAGNN	0.41	0.41	0.41	0.31	0.36	0.33	0.73	0.77	0.75
HSGNN	0.66	0.64	0.65	0.75	0.75	0.75	0.69	0.48	0.59
GraphGPS	0.49	0.56	0.44	-	-	-	0.23	0.28	0.17
Graphormer	0.90	0.90	0.90	0.69	0.57	0.63	0.91	0.90	0.90
GRAM	0.71	0.73	0.72	0.79	0.81	0.78	0.64	0.54	0.61
MDP-GRL	0.74	0.79	0.76	0.78	0.81	0.78	0.81	0.72	0.76
EHR2HG	0.16	0.20	0.18	0.16	0.20	0.17	-	-	-
RGCN	0.92	0.90	0.90	0.94	0.95	0.95	0.99	0.99	0.99
GIE	0.78	0.56	0.73	0.78	0.58	0.75	0.82	0.82	0.82
HGCGE	0.71	0.57	0.62	0.70	0.70	0.69	0.65	0.65	0.64
AttH	0.75	0.62	0.75	0.83	0.38	0.59	0.77	0.61	0.68
HNMP-EHR (ours)	0.98	0.99	0.99	0.99	0.99	0.99	0.97	0.98	0.98

The performance metrics for the MIMIC III dataset, as shown in Table 4, demonstrate the varying effectiveness of graph-based methods across single diagnosis, lab, and medication prediction tasks. HNMP-EHR achieves F1 scores in the range 0.92-0.98, comparable with EHR2HG and MAGNN which achieve perfect F1 scores in diagnosis and lab tasks. Other metapath based models including Metapath2Vec under-perform in diagnosis (f1 score 0.35). On a single task, other models including MDP-GRL with f1 score of 0.48, on medication prediction also exhibits below the threshold for practical applications.

Similarly for the MIMIC IV dataset, as shown in Table 4, the single-task performance of HNMP-EHR ranges from 0.88 to 0.99, closely matching EHR2HG and MAGNN (highest) in diagnosis and lab tasks. Metapath2Vec as a meta-path only method excels in lab and medication predictions but underperforms in diagnosis outcome prediction (f1 score of 0.74). Global attention models show a notably low diagnosis score, including GRAM (with f1 score of 0.21), while other general purpose models including RGCN (f1 score 0.58 for Lab tests) also fail to achieve good performance across the board. These results once again underscore HNMP-EHR's consistent and robust performance for practical settings.

Lastly, the performance on eICU dataset as presented in Table 4, reveals the effectiveness of various methods across diagnosis, lab, and medication prediction tasks for multi-center data. MAGNN as a meta-path based graph neural network excels in diagnosis and lab test (both f1 score 1.00) tasts, but underperforms in medication prediction (with f1 score 0.23). Metapath2Vec as another instance of meta-path approach performs well in two tasks, namely lab and medication with perfect f1 scores, but lags in diagnosis outcome prediction (f1 score 0.68). HNMP-EHR (PE) scores in the 0.97-0.98 range for two tasks and 0.78 points in F1 score for medication recommendation. It's performance closely matches RGCN (E), suggesting the structure-less nature of these tasks using eICU datasets.

Table 4: Single-task performance of models on MIMIC III, MIMIV IV, and eICU datasets in terms of Precision, Recall, and F1 score for Diagnosis, Lab Test, and Medication prediction.

Method	MIMIC III								
	Diagnosis			Lab			Medication		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Metapath2Vec	1.00	0.21	0.35	1.00	0.99	0.99	1.00	0.99	0.99
MAGNN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.54	0.70
HSGNN	1.00	0.91	0.95	1.00	0.96	0.97	1.00	0.54	0.69
GraphGPS	1.00	0.52	0.68	1.00	0.64	0.77	1.00	0.60	0.72
Graphormer	1.00	0.97	0.98	1.00	0.55	0.68	1.00	0.99	0.99
GRAM	-	-	-	1.00	0.99	0.99	1.00	0.83	0.86
MDP-GRL	1.00	0.78	0.87	1.00	0.85	0.92	1.00	0.31	0.48
EHR2HG	1.00	1.00	1.00	1.00	1.00	1.00	-	-	-
RGCN	1.00	0.87	0.93	1.00	0.84	0.90	1.00	0.90	0.94
GIE	1.00	0.92	0.96	1.00	0.99	0.99	1.00	0.84	0.91
HGCGE	0.04	0.39	0.07	0.57	0.41	0.47	0.39	0.38	0.38
AttH	1.00	0.95	0.98	1.00	0.99	0.99	1.00	0.76	0.87
HNMP-EHR (ours)	1.00	0.87	0.92	1.00	0.93	0.96	1.00	0.97	0.98
Method	**MIMIC IV**								
	Diagnosis			Lab			Medication		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Metapath2Vec	1.00	0.59	0.74	1.00	1.00	1.00	1.00	1.00	1.00
MAGNN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.68	0.81
HSGNN	1.00	0.98	0.99	1.00	0.64	0.78	1.00	1.00	1.00
GraphGPS	0.88	0.53	0.65	1.00	0.67	0.79	0.88	0.53	0.64
Graphormer	1.00	0.99	0.99	1.00	0.53	0.65	1.00	0.90	0.93
GRAM	1.00	0.12	0.21	1.00	0.68	0.72	1.00	1.00	1.00
MDP-GRL	1.00	0.84	0.91	1.00	0.76	0.86	1.00	0.49	0.65
EHR2HG	1.00	1.00	1.00	1.00	1.00	1.00	-	-	-
RGCN	1.00	0.85	0.91	1.00	0.45	0.58	1.00	0.94	0.97
GIE	1.00	0.99	0.99	1.00	0.66	0.79	1.00	0.86	0.92
HGCGE	0.26	0.35	0.30	0.34	0.48	0.32	0.47	0.42	0.44
AttH	1.00	0.73	0.84	1.00	0.91	0.95	1.00	0.77	0.87
HNMP-EHR (ours)	1.00	0.99	0.99	1.00	0.81	0.88	1.00	0.99	0.99
Method	**eICU**								
	Diagnosis			Lab			Medication		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Metapath2Vec	1.00	0.52	0.68	1.00	1.00	1.00	1.00	1.00	1.00
MAGNN	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.16	0.23
HSGNN	0.70	0.13	0.19	1.00	0.99	0.99	0.30	0.15	0.19
GraphGPS	1.00	0.5	0.66	1.00	0.57	0.72	1.00	0.52	0.67
Graphormer	1.00	0.28	0.43	1.00	0.76	0.86	1.00	0.73	0.83
GRAM	1.00	0.39	0.56	1.00	0.52	0.67	1.00	0.52	0.68
MDP-GRL	1.00	0.78	0.87	1.00	0.52	0.68	1.00	0.77	0.87
EHR2HG	-	-	-	-	-	-	-	-	-
RGCN	1.00	0.93	0.96	1.00	0.98	0.99	1.00	0.66	0.78
GIE	1.00	0.83	0.90	1.00	0.99	0.99	1.00	0.92	0.96
HGCGE	0.20	0.45	0.28	0.76	0.55	0.64	0.22	0.44	0.30
AttH	1.00	0.67	0.80	1.00	0.99	0.99	1.00	0.85	0.92
HNMP-EHR (ours)	1.00	0.97	0.98	1.00	0.95	0.97	1.00	0.73	0.78

D PRELIMINARIES AND NOTATION

D.1 GRAPH NEURAL NETWORKS ON PATIENT AND ONTOLOGY GRAPHS

Let $\mathcal{G}_p = (\mathcal{C}_p, \mathcal{R}_p)$ denote the *patient graph*, where $\mathcal{C}_p$ is the set of patient codes (observations $\mathcal{O}_p$, diagnoses $\mathcal{D}_p$, and medications $\mathcal{M}_p$), and $\mathcal{R}_p$ represents intra-visit relationships among these entities. Similarly, let $\mathcal{G}_q = (\mathcal{C}_q, \mathcal{R}_q)$ denote the *ontology graph*, where $\mathcal{C}_q$ contains ontology codes and $\mathcal{R}_q$ captures semantic relations between them. For a node $p_i \in \mathcal{C}_p$, let $\mathcal{N}_i^r = \{p_j \in \mathcal{C}_p : (p_i, p_j) \in \mathcal{R}_p\}$ denote the set of neighbors connected via relation type r. Similarly, $\mathcal{N}_i^r$ for $q_i \in \mathcal{C}_q$ is defined analogously. Let $c_{i,r}$ be a normalization constant based on the degree of node i under relation r. The standard GCN message passing rule at layer l for a patient node p_i is then defined as:

$$m_{\to p_i}^{(l)} = \sum_{r \in \mathcal{R}_p} \sum_{p_j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_{p_j}^{(l-1)},$$

and:

$$h_{p_i}^{(l)} = \sigma\Big(W_0^{(l)} h_{p_i}^{(l-1)} + m_{\to p_i}^{(l)}\Big)$$

where $W_r^{(l)}$ and $W_0^{(l)}$ are relation-specific and self-transformation matrices, respectively, and $\sigma(\cdot)$ is a nonlinear activation function. A similar message passing formulation applies to ontology nodes $q_i \in \mathcal{C}_q$ using $\mathcal{R}_q$.

D.2 HYPERBOLIC GRAPH NEURAL NETWORKS

While conventional GCNs operate in Euclidean space ($\kappa_1 = 0$), graphs with hierarchical or tree-like structures, such as ontologies, are better modeled in hyperbolic space ($\kappa_2 = 1$). Hyperbolic Graph Neural Networks (HGNNs) extend the message passing paradigm by replacing Euclidean operations with their hyperbolic counterparts in the Poincaré ball $\mathbf{B}_c^d$.

Lifting and projection. Given an ontology node embedding $h_{q_i}^{(l-1)}$ in $\mathbf{B}_c^d$, we first map it to the tangent space $T_0\mathbf{B}_c^d$ via the logarithmic map $\log_0^B(\cdot)$, apply transformations in the Euclidean tangent space, and map it back to the manifold using the exponential map $\exp_0^B(\cdot)$:

$$\tilde{h}_{q_i}^{(l-1)} = \log_0^B(h_{q_i}^{(l-1)}), \quad h_{q_i}^{(l)} = \exp_0^B(\tilde{h}_{q_i}^{(l)}).$$

Hyperbolic feature transformation. Feature transformation in hyperbolic space uses Möbius matrix-vector multiplication. At layer l, the transformed representation is:

$$h_{q_i}^{(l)} = \exp_0^B\left(W_r^{(l)} \log_0^B(h_{q_i}^{(l-1)})\right)$$

where $W_r^{(l)}$ is the trainable relation-specific weight matrix in the tangent space.

Hyperbolic neighborhood aggregation. Aggregation in hyperbolic space is performed in the tangent space to enable linear operations:

$$h_{q_i}^{(l)} = \exp_0^B\left(\sigma\Big(\log_0^B(h_{q_i}^{(l)}) + \sum_{q_j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} \log_0^B(h_{q_j}^{(l-1)})\Big)\right)$$

This formulation ensures that the aggregated representation respects the manifold geometry while incorporating information from neighboring nodes.

Cross-graph alignment. To bridge the patient and ontology graphs, cross-graph messages $m'^{(l)}_{\to p_i}$ and $m'^{(l)}_{\to q_i}$ are exchanged between aligned anchors $(p_i, q_i) \in \mathcal{A}$:

$$m'^{(l)}_{\to p_i} = W_s^{(l)} \log_0^B(h_{q_i}^{(l-1)}), \quad m'^{(l)}_{\to q_i} = W_s^{(l)} \log_0^B(h_{p_i}^{(l-1)})$$

where $W_s^{(l)}$ is the alignment transformation matrix. These cross-graph signals are incorporated into the update of each graph's node representation.

Riemannian optimization. Training parameters $\theta \in \{W_r^{(l)}, W_0^{(l)}, W_s^{(l)}\}$ in hyperbolic space requires Riemannian optimization. The Euclidean gradient $\nabla_E \mathcal{L}(\theta)$ is projected to the tangent space:

$$\nabla_R \mathcal{L}(\theta) = \text{Proj}_{T_\theta \mathbf{B}_c^d}\big(\nabla_E \mathcal{L}(\theta)\big)$$

and parameters are updated using the exponential map:

$$\theta^{(t+1)} = \exp_\theta^B\left(-\eta \, \nabla_R \mathcal{L}(\theta)\right)$$

ensuring that updates remain on the manifold throughout training.

Table 5: Summary of Notations

Notation	Description
$\mathcal{G}_p$	Patient graph composed of visits, observations, diagnoses, and medications
$\mathcal{G}_q$	Ontology graph representing structured hierarchical clinical knowledge
$\mathcal{O}_p, \mathcal{D}_p, \mathcal{M}_p$	Sets of observations, diagnoses, and medications for patient data
$\mathcal{O}_q, \mathcal{D}_q, \mathcal{M}_q$	Sets of ontology observations, diagnoses, and medications
$\mathcal{C}_p$	Union of all patient codes: $\mathcal{O}_p \cup \mathcal{D}_p \cup \mathcal{M}_p$
$\mathcal{C}_q$	Union of all ontology codes: $\mathcal{O}_q \cup \mathcal{D}_q \cup \mathcal{M}_q$
v_p^t	The t-th visit of patient p, containing a subset of codes from $\mathcal{C}_p$
$\mathcal{A}$	Set of anchor pairs linking patient and ontology nodes.
$\mathcal{A}_O, \mathcal{A}_D, \mathcal{A}_M$	Subsets of anchors for observations, diagnoses, and medications
$\mathcal{R}_p$	Relation set encoding intra-visit relationships among patient entities
$\mathcal{R}_q$	Relation set encoding semantic relations within the ontology
$\mathcal{N}_i^r$	Set of r-type neighbors of node i
$c_{i,r}$	Normalization constant based on degree of node i under relation r
p_i, q_i	Nodes in the patient graph (p_i) and ontology graph (q_i)
$h_{p_i}^{(l)}, h_{q_i}^{(l)}$	Representation of node p_i or q_i at layer l
$m_{\to p_i}^{(l)}, m_{\circlearrowleft p_i}^{(l)}$	Neighbor and self-messages for patient node p_i at layer l
$m_{\to q_i}^{(l)}, m_{\circlearrowleft q_i}^{(l)}$	Neighbor and self-messages for ontology node q_i at layer l
$m_{\to p_i}^{\prime(l)}, m_{\to q_i}^{\prime(l)}$	Cross-graph alignment messages from ontology to patient and vice versa
$W_r^{(l)}, W_0^{(l)}, W_s^{(l)}$	Relation-specific, self, and alignment transformation matrices
κ_1	Curvature parameters for patient space ($\kappa_1 = 0$, Euclidean)
κ_2	Curvature parameters for ontology space ($\kappa_2 = 1$, hyperbolic)
$\mathbf{B}_c^d$	d-dimensional Poincaré ball of radius c, used for hyperbolic embeddings
$\log_x^B(\cdot), \exp_x^B(\cdot)$	Logarithmic and exponential maps in the Poincaré ball
$\oplus_c$	Möbius addition operator in hyperbolic space
$f(s, r, t)$	DistMult scoring function for triples in patient graph
$\varphi(s, r, t)$	Möbius scoring function for triples in ontology graph
$\mathcal{T}, \mathcal{T}_{\text{neg}}$	Sets of positive and negative triples for link prediction
γ	Margin for ranking loss (set to 1.0)
$\mathbf{h}_p^t$	Pooled visit-level embedding for patient p at visit t
$\hat{y}_p^t$	Predicted survival probability for patient p at visit t
y_p^t	Ground-truth survival label for patient p at visit t
$\mathcal{L}_{\text{link}}^p, \mathcal{L}_{\text{link}}^q$	Link prediction losses for patient and ontology graphs
$\mathcal{L}_{\text{visit}}$	Visit-level survival prediction loss
$\mathcal{L}_{\text{concurrent}}$	Joint training objective combining patient, ontology, and visit losses

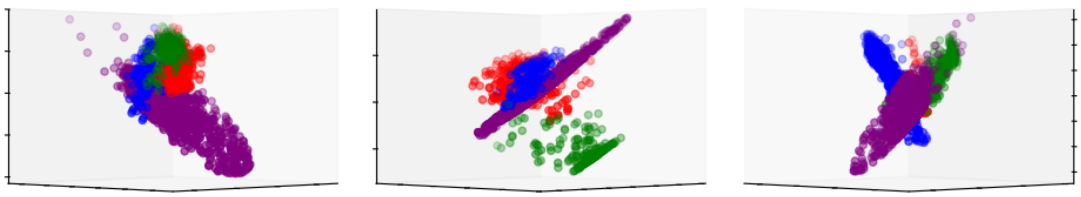


Figure 10: 3D Visualization of dual-space embeddings (d=3) generated by HNMP and displayed from different views, including Diagnosis (red), Medication (blue), Lab Test (green) nodes in Euclidean (E) space, and Ontology (purple) nodes in Poincare (P) space.