Concurrent Clinical Predictions via Dual-Space Message Passing and Ontology Alignment

Shervin Mehryar
Department of Advanced Computing Sciences
Maastricht University
Maastricht, Netherlands
shervin.mehryar@maastrichtuniversity.nl

Michel Dumontier
Department of Advanced Computing Sciences
Maastricht University
Maastricht, Netherlands
michel.dumontier@maastrichtuniversity.nl

***Abstract*—The growing availability of electronic health records (EHR) presents increasing opportunities for data-driven applications in healthcare. These records, however, are represented using disparate coding systems and often suffer from fragmentation and structural inconsistencies. Such data fragmentation can lead to substantial performance gaps across different predictive tasks. In this work, a framework is proposed in order to mitigate these artefacts and enable multi-task predictions. To address performance gaps, we introduce a concurrent loss that guides the learning process over fragmented entities. To address structural misalignment, we propose a neural message-passing algorithm that enables the integration of ontological information. Experiments conducted on cardiovascular disease (CVD) patients demonstrate that the proposed system overcomes prediction gaps across multiple clinical tasks, including laboratory tests, diagnoses, medications, and mortality prediction, performed concurrently. The results highlight the potential of ontology-aligned representation learning for high-precision decision-making in healthcare systems.**

***Index Terms*—Electronic Health Records, Neuro-symbolic AI, Graph Neural Networks, Ontology Alignment, CVD, Concurrent Prediction**

I. 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. The increasing availability of clinical datasets has enabled predictive models to support decision-making tasks such as diagnosis prediction [1], medication recommendation [2], and laboratory testing [3], with the goal of improving patient treatment. However, the predictive performance of data-driven models can be substantially impaired by errors and inconsistencies in patient records. In particular, legacy EHR systems rely on multiple medical terminologies to encode patient data, which differ both semantically and structurally. These diverse terminologies introduce significant data fragmentation and can lead to adverse outcomes, such as prescription or procedural oversights [4].

Ontology alignment is a technique that addresses semantic inconsistencies in EHR by mapping diverse data elements to a unifying resource, i.e. a reference ontology [5]. Prior work has demonstrated that semantic alignment is a critical step toward resolving ambiguous representations and enabling robust predictive modeling [6]. Moreover, biomedical ontologies encode rich domain-specific and hierarchical knowledge that can be leveraged to improve precision in downstream prediction tasks [7]. Despite these advantages, extending ontology alignment to support concurrent predictions while maintaining the high level of precision required for clinical adoption, remains an open research problem [3].

Recent advances in Graph Neural Networks (GNN) have provided a powerful framework for modeling the heterogeneous structure of EHR for downstream tasks via node and link prediction [8]. GNNs in particular excel at capturing local and global dependencies across diverse biomedical concepts, medical codes, and patient admission records, through non-linear neural operators and iterative message passing. As a neurosymbolic paradigm, GNN-based approaches can facilitate a wide range of clinically relevant predictive as well as alignment tasks with high accuracies [9].

In this paper, we propose a framework that enables multiple high-precision clinical tasks by integrating ontological knowledge with patient information through neuro-symbolic 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 [10]. 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 message passing for concurrent tasks is a novel approach.

The rest of this paper is organized as follows. We provide a description of different clinical tasks from the literature along with existing paradigms in GNN-based predictive modeling in Section II. The proposed framework using a dual-space message passing mechanism is presented in Section III, including objective functions and algorithmic specifications. We run experiments using the records of patients with Cardiovascular Disease (CVD) from two real-world datasets, described in Section IV-A. Concurrent performance results are presented in IV-B and the contribution of ontology alignment to multi-task performance at scale is presented in IV-C.

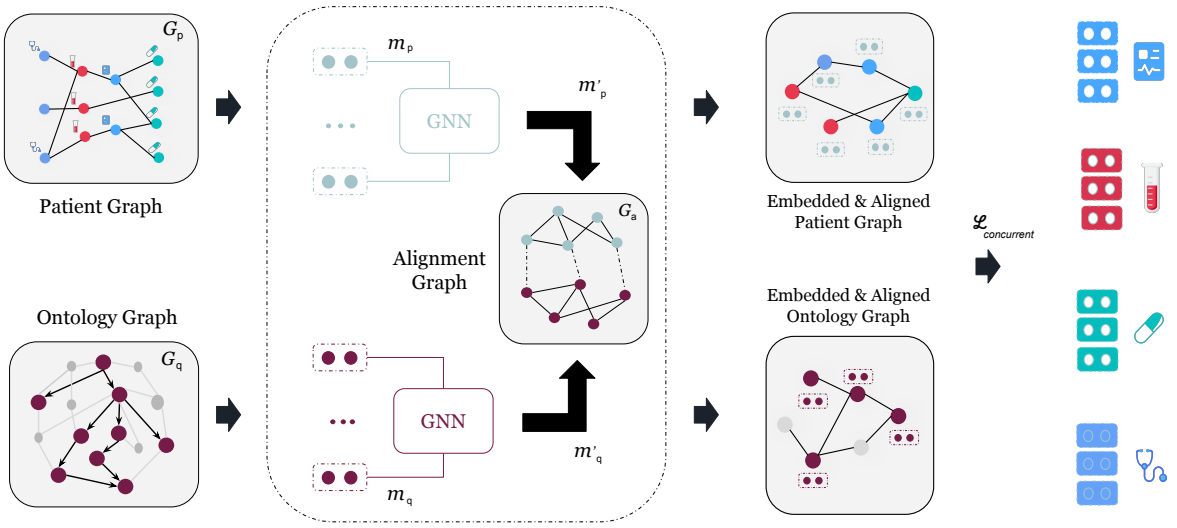


Fig. 1: The alignment and prediction framework with neural message passing for concurrent clinical tasks.

II. Related Work

A summary of the state-of-the-art research for relevant applications based on heterogenuous graph representation learning are provided in this section. For the **diagnosis** task, the proposed model in [1] constructs patient record graphs from medical codes and external knowledge bases, using a neural graph encoder to generate embeddings for patients and diseases, enabling inductive inference for new patients. The approach in [7] casts this objective as a link prediction task using relational graph neural networks. For the single task of diagnosis code prediction and through alignment with a reference ontology, it achieves high precision levels required for adoption in practical settings. For **medication** recommendations, the approach in [11] leverages drug features from an external knowledge base to complement patient information, thus improving baseline precision, recall, and f1 metrics. The method in [3] forms entities and encounters/visits as a bipartite graph and applies message passing for **laboratory** observation prediction, jointly with medication recommendation. In contrast, **mortality** prediction is performed at visit level as node prediction and has been the subject of research as a standalone task [12]. In our work, we propose a mixture of link and node prediction objectives that support these applications in a multi-task setting [13].

III. 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 a 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 by unique identifiers (i.e. code) 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 embedding space to learn the representations over the EHR data and a different embedding space to learn the ontological representations. The dual representations are aligned using neural messages passed between a set of anchoring nodes [14].

More formally, let $\mathcal{O}_p$ denote the set of medical observations, $\mathcal{D}_p$ denote diagnoses, and $\mathcal{M}_p$ denote medications 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. $v_p^t \in \mathcal{C}_p$). In this work, the inter-visit relations for a given patient are not considered and the superscript t is dropped when appropriate. Conversely, concepts from a reference ontology $\mathcal{C}_q$ are organized as a collection of ontology observations $\mathcal{O}_q$, diagnoses $\mathcal{D}_q$, and medication codes $\mathcal{M}_q$, 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 ontological subgraphs (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 (i.e. visit to diagnosis, observation, or medication codes). In parallel, a separate collection of edges $\mathcal{R}_q$ is taken to encode semantic relations within the ontology (i.e. parent-child relation). Lastly, anchor representations are aligned through a dedicated relation (i.e. equivalent) and message computations.

These intra-visit, intra-ontology, and cross-anchor message-passing operations form the basis of the proposed representation learning framework, as detailed out in the next section.

A. Patient Graph Message Computations

The patient graph $\mathcal{G}_p$ includes information related to patient visits encoded by $\mathcal{C}_p$. The representations for patient visits are learned using 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: $m_{p_i}^{(l)} = m_{\circlearrowleft p_i}^{(l)} + m_{\rightarrow p_i}^{(l)}$, 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:

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

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, $\mathbf{W}_r^{(l)}$ 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:

$$m_{\circlearrowleft p_i}^{(l)} = \mathbf{W}_0^{(l)} h_{p_i}^{(l)}, \tag{2}$$

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 improves model interpretability and modularity while preserving the inductive relational biases critical for downstream clinical tasks (as shown in Figure 2a).

B. 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 neural message passing over an L layer graph neural network. 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: $m_{q_i}^{(l)} = m_{\circlearrowleft q_i}^{(l)} + m_{\rightarrow q_i}^{(l)}$, 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:

$$m_{\rightarrow q_i}^{(l)} = \sum_{r \in \mathcal{R}_q} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} \mathbf{W}_r^{(l)} h_{q_j}^{(l)}, \tag{3}$$

where $\mathcal{R}_q$ are relation types, $\mathcal{N}_i^r$ is the set of 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:

$$m_{\circlearrowleft q_i}^{(l)} = \mathbf{W}_0^{(l)} h_{q_i}^{(l)}, \tag{4}$$

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 sigmoid function. By explicitly separating self-information and neighbor aggregation, relational as well as hierarchical information are preserved (as shown in Figure 2c).

C. Alignment Graph Message Computations

The alignment graph $\mathcal{G}_a$ contains anchor pairs between the patient graph $\mathcal{G}_p$ and the ontology graph $\mathcal{G}_q$. Ontological knowledge transfer is achieved through inter-graph message aggregation and transformation as follows. For a node p_i in the patient graph, the incoming message aggregation from its ontology neighbors at layer l is computed by: $m_{p_i}^{\prime(l)} = m_{\circlearrowleft p_i}^{(l)} + m_{\rightarrow p_i}^{\prime(l)}$, involving self-information message $m_{\circlearrowleft p_i}^{(l)}$ defined as before and the inter-graph message term given by:

$$m_{\rightarrow p_i}^{\prime(l)} = \sum_{j \in \mathcal{N}_i^{'}} \frac{1}{c_i} \mathbf{W}_s^{(l)} h_{q_j}^{(l)}, \tag{5}$$

where $\mathbf{W}_s^{(l)}$ 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: $m_{q_i}^{\prime(l)} = m_{\circlearrowleft q_i}^{(l)} + m_{\rightarrow q_i}^{\prime(l)}$, involving self-information message $m_{\circlearrowleft q_i}^{(l)}$ defined as before and the inter-graph message term given by:

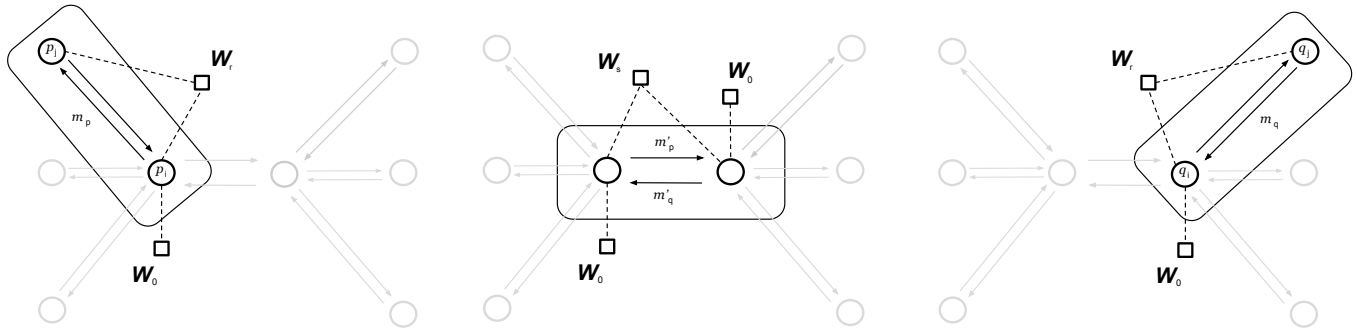
$$m_{\rightarrow q_i}^{\prime(l)} = \sum_{j \in \mathcal{N}_i^{'}} \frac{1}{c_i} \mathbf{W}_s^{(l)} h_{p_j}^{(l)}, \tag{6}$$

where $\mathbf{W}_s^{(l)}$ denotes the alignment weight 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 $h_{p_i}^{(l+1)} = \sigma\left(m_{p_i}^{\prime(l)}\right)$ and $h_{q_i}^{(l+1)} = \sigma\left(m_{q_i}^{\prime(l)}\right)$, where σ denotes a pointwise nonlinearity such as the sigmoid function. These cross-graph updates ensure bidirectional alignment by exchanging mutual information between anchor nodes (as shown in Figure 2b).

D. Concurrent Predictions

In the remaining part of this section, we introduce an objective function to combine task-specific and ontological features. In particular, a mixture of link prediction tasks between each visit node and nodes representing clinical codes for diagnosis, medication, and observation are considered [13]. Mortality is determined based on a binary classification on a patient's visit representation as a node prediction task [15].

Link Prediction: over the patient graph $\mathcal{G}_p$, we employ a DistMult decoder [16] to score relations between visits and



(a) Patient messages between node p_i & node p_j.

(b) Alignment between patient p_i & ontology node q_i.

(c) Ontology messages between node q_i & node q_j.

Fig. 2: Message passing diagrams for nodes shown as circles and trainable parameters shown as squares.

clinical codes as triples (s, r, t) using a margin-based ranking loss defined by:

$$\mathcal{L}^p_{\text{link}} = \frac{1}{|\mathcal{T}|} \sum_{\substack{(s,r,t)\in\mathcal{T} \\ (s',r,t')\in\mathcal{T}_{\text{neg}}}} \max(0, \gamma - f(s,r,t) + f(s',r,t')), \tag{7}$$

with the bi-linear inner product given by $f(s,r,t) = \langle \mathbf{h}_s, \mathbf{h}_r, \mathbf{h}_t \rangle$. Here, $\mathbf{h}_s$ and $\mathbf{h}_t$ denote node embeddings for patient visits $v_p \in \mathcal{V}_p$ as source and clinical codes $\mathcal{C}_p$ as tail, with $\mathbf{h}_r$ embedding their specific relation. 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 samples. Negative edges are generated at 1-to-1 positive to negative ratio, by corrupting either the source or tail node uniformly at random. During evaluation, a link is predicted as positive (or negative) if its score is above (or below) a decision threshold λ (set to 0.5). Standard precision, recall, and f1-scores are reported on positive (existing) and negative (non-existing) edges accordingly. A similar loss term $\mathcal{L}^q_{\text{link}}$ is utilized for the ontology graph $\mathcal{G}_q$.

Node Prediction: for a visit $v_p \in \mathcal{V}_p$ given a patient p with a set of corresponding medical codes $\mathcal{C}_p$, we obtain a visit-level representation by: $\mathbf{h}_p = g(\{\mathbf{h}_c : c \in \mathcal{C}_p\})$, where $g(.)$ is a pooling operator. The pooled embedding $\mathbf{h}_p$ is mapped to a binary value for patient survival prediction at visit level by: $\hat{y}_p = \sigma(\mathbf{w}^\top \mathbf{h}_p + b)$, where $\sigma(\cdot)$ is the logistic sigmoid, $\mathbf{w}$ and b are learnable parameters, and $\hat{y}_p \in [0, 1]$ is the predicted probability of survival. The model is trained with binary cross-entropy loss:

$$\mathcal{L}_{\text{visit}} = -\frac{1}{|\mathcal{D}_S|} \sum_{p\in\mathcal{D}_S} \Big(y_p \log \hat{y}_p + (1 - y_p) \log(1 - \hat{y}_p) \Big), \tag{8}$$

where $y_p \in \{0, 1\}$ is the ground-truth survival label for patient p's visit and $\mathcal{D}_S$ is the set of patient visits used for training. This formulation encourages the model to embed visit-level embeddings towards survival outcome prediction. In this paper, the long term relation between visits is not considered and mortality predictions are made based on individual visits independently.

Algorithm 1 Ontology-Aligned Message Passing

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

for each patient node p_i in parallel **do** ▷ patient updates
  Compute self-message $m^{(l)}_{\circlearrowleft p_i}$ by Eq. (2)
  Compute relational message $m^{(l)}_{\to p_i}$ by Eq. (1)
  Aggregate message: $m^{(l)}_{p_i} \leftarrow m^{(l)}_{\circlearrowleft p_i} + m^{(l)}_{\to p_i}$
end for

for each ontology node q_i in parallel **do** ▷ ontology updates
  Compute self-message $m^{(l)}_{\circlearrowleft q_i}$ by Eq. (4)
  Compute relational message $m^{(l)}_{\to q_i}$ by Eq. (3)
  Aggregate message: $m^{(l)}_{q_i} \leftarrow m^{(l)}_{\circlearrowleft q_i} + m^{(l)}_{\to q_i}$
end for

for each pair $(p_i, q_j) \in \mathcal{A}$ in parallel **do** ▷ alignment updates
  Compute patient cross-message $m'^{(l)}_{\to p_i}$ by Eq. (5)
  Aggregate message: $m'^{(l)}_{p_i} \leftarrow m^{(l)}_{\circlearrowleft p_i} + m'^{(l)}_{\to p_i}$
  Compute ontology cross-message $m'^{(l)}_{\to q_j}$ by Eq. (6)
  Aggregate message: $m'^{(l)}_{q_j} \leftarrow m^{(l)}_{\circlearrowleft q_j} + m'^{(l)}_{\to q_j}$
end for

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

$$\mathcal{L}_{\text{concurrent}} = \mathcal{L}^p_{\text{link}} + \mathcal{L}^q_{\text{link}} + \mathcal{L}_{\text{visit}}. \tag{9}$$

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.

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 $\mathcal{A}$. For each patient graph node p_i, the self-messages are computed based on Eq. 2 and relational messages via Eq. 1, aggregated as m_{p_i}. Similarly, for ontology graph nodes q_i, it computes and aggregates self (via Eq. 4) and relational messages (via Eq. 3) into m_{q_i}. 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 Eq. 5 and Eq. 6).

IV. Experimentation & Results

A. EHR Datasets & Reference Ontology

We use publicly available repositories containing tabular data to generate patient graphs and limit the scope to records from patients that are hospitalized for Cardiovascular Disease (CVD). The selected records are grouped by visits and structured around three core concepts, namely Diagnosis, Medication, and Lab Observations, similar to [17]. The two datasets are:

MIMIC III: containing data associated with admissions to critical care units, including ICD-9 coding for diagnosis and procedures, LOINC for lab observations, and NDC for drugs/medication [18].

eICU: a multi-center dataset containing high-granularity admission data, including ICD-9 coding for diagnoses, LOINC identifiers for lab observations, and RxNorm terminology for medication [19].

The CVD admissions selected for experimentation include diagnosis ICD-9 code in the range 410-430 which categorize various forms and severities of heart failure. We identify and include admissions for those patients that have at least one of the above ICD codes associated with them. Three relation types are created, namely 'has lab code', 'has diagnosis code', and 'has drug code' to link each visit to respective entities in the patient graph $\mathcal{G}_p$. We note that a patient's visit v_p may include entities from one or multiple coding systems which we reconcile through ontology alignment.

For domain specific alignment of entities, we utilize the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). SNOMED is a comprehensive ontology used to encode clinical information, including diagnosis, medication, and observations in a single, unifying resource. Mappings from EHR codification (e.g. ICD, LOINC, NDC, etc) to SNOMED concepts are publically available as reference sets[1]. Alignments along with their transitive closures are added via the 'equivalent' and 'subclass' predicates, respectively, to $\mathcal{G}_a$.

After processing the CVD visits and generating one-to-one and one-to-many maps for available entities, the alignment graph includes 570 diagnosis, 255 drug, and 458 lab codes in the case of MIMIC III dataset (1,360 unique nodes including transitive closures), as well as, 1,580 diagnosis, 761 drug, and 495 lab codes in the case of eICU dataset (13,999 unique nodes including transitive closures), as shown in Table I.

B. Concurrent Performance

We report on the performance of the proposed framework for Concurrent Clinical Prediction with Ontology-Aligned Message Passing (CCP-OAMP) using precision, recall, and f1-scores. In this set of experiments, training and evaluation were repeated across 10 independent runs, using Algorithm 1 with $L = 2$ layers to optimize the objective function in Eq. 9. In each run, a subset of 100 random patient visits (i.e. $|\mathcal{D}_S| = 100$) is selected for training and evaluation with node and edge stratification.

For comparison, we consider four categories of baseline models ranging from graph-based to EHR-specific methods. The first category consists of models that rely on Meta-paths (M) for heterogeneous structure learning, including Meta2vec [20], MAGNN [21], and HSGNN [22]. The second category of baseline models include Graphormer [23] and GraphGPS [24] which encode graph-structured information (G) using a Transformer architecture. The third category of models consists of hierarchical learning on a Hyperbolic manifold (H), including HGCGE [25] and AttH [26]. In the fourth category, we consider models developed specifically for clinical tasks (C), including GRAM [27] and MDP-GRL [28]. All models were trained with default parameters on 80% of data and tested

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

TABLE I: Statistics for CVD patient graph $\mathcal{G}_p$, ontology graph $\mathcal{G}_q$, and alignment graph $\mathcal{G}_a$ using MIMIC III and eICU datasets, in terms of the number of unique codes for diagnosis $\mathcal{D}_p$, medication $\mathcal{M}_p$, and lab observations $\mathcal{O}_p$.

	MIMIC III		**eICU**	
	$\mathcal{G}_p$	$\mathcal{G}_q$	$\mathcal{G}_p$	$\mathcal{G}_q$
$\lvert\mathcal{D}_p\rvert$	149	570	399	1580
$\lvert\mathcal{M}_p\rvert$	969	255	191	761
$\lvert\mathcal{O}_p\rvert$	327	458	188	495
$\lvert\mathcal{G}_a\rvert$	–	1360	-	13999

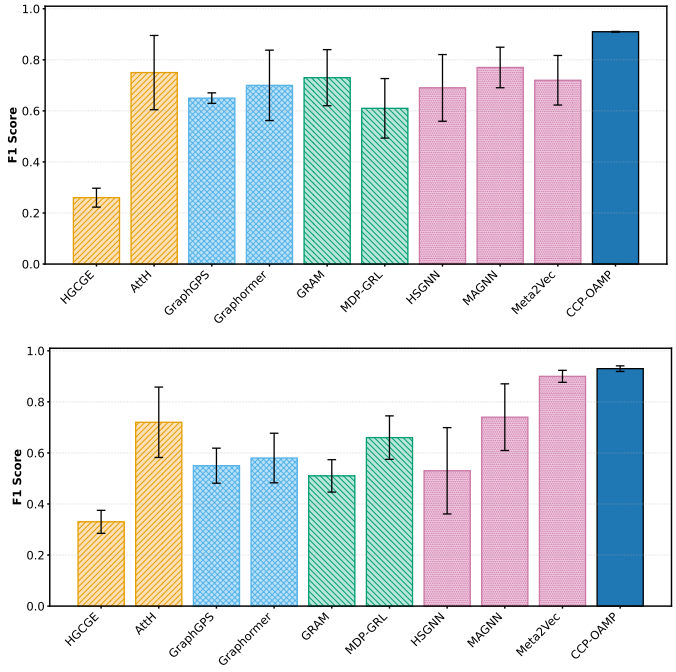


Fig. 3: Comparison of mean and varience in F1 Score across tasks between baseline models and proposed CCP-OAMP (dark blue), over 10 independent runs using 100 patient visits from MIMIC-III (top) and eICU (bottom) datasets. Model categories are: Meta-path Models (M) in purple, Graph Transformers (G) in light blue, Clinical specific (C) in green, and Hyperbolic (H) in yellow.

TABLE II: Concurrent performance by the proposed CCP-OAMP model on graphs with 100 patient visits from MIMIC III and eICU datasets in terms of Precision, Recall, and F1 Score for Link Prediction (Diagnosis/Observations/Medication) and Node Prediction (Mortality).

Dataset	Link Prediction			Node Prediction		
	Precision	Recall	F1	Precision	Recall	F1
MIMIC III	1.00	0.87	0.92	0.92	0.90	0.90
eICU	1.00	0.85	0.91	0.99	0.99	0.99

on 20% of data held separately.

As shown in Figure 3, the proposed CCP-OAMP model achieves consistent f1-scores (over 0.90 points) across diverse tasks and on both MIMIC III and eICU patient graphs with SNOMED alignments. In category M models, performance in this experiment setting remains below 0.90 points. Among these models, Meta2Vec and MAGNN exhibit relatively better performance due to the utilization of meta-paths which guide and transfer heterogeneous information. Models in category C exhibit performance below 0.80 points in f1-score. GRAM (developed for visit-level prediction incorporating code hierarchies) is observed to under-perform in multi-task settings. MDP-GRL (developed for multi-label prediction) also exhibits under-performance which is attributed to fragmentations in codification. Among models in category G, the concurrent performance falls below 0.60 points in large part due to the diversity of tasks that require multiple levels of structure-aware learning. For instance, GraphGPS (with modular components for structural encoding of data given a single task) lacks an explicit architecture to deduce interactions among tasks [29]. The Hyperbolic (H) models including AttH and HGCGE, despite having explicit hierarchical encoding, fail to capture chain-like data structures common specially in clinical graphs. As evident, baseline models that treat input graphs as structure-less or non-fragmented underachieve in comparison.

The proposed CCP-OAMP model separates embedding spaces to capture mixed interactions (i.e. a dual-space representation) and mitigates data fragmentation for concurrent predictions with hierarchical alignment. As a result, CCP-OAMP achieves consistently high performance across tasks and on diverse patient graphs, further detailed in Table II. Over patient graphs from MIMIC III, the model achieves f1-score of 0.92 on diagnosis, observations, and medication related link prediction tasks, as well as f1-score of 0.90 on mortality node prediction task. Over patient graphs from eICU, f1-scores of 0.91 and 0.99 are achieved across different link and node prediction tasks. Notably, CCP-OAMP attains high precision (above 0.90 in all cases), a key requirement for adoption in practical clinical settings.

C. Effect of Ontology-aligned Message Passing

We conduct a second set of experiments to showcase the importance of information exchange through ontology-alignment for multi-task purposes. More specifically, a number of patient graphs are generated with increasing sizes from 50 to 500 visits in increments of 50 (i.e. $|\mathcal{D}_S| = 50$ to $|\mathcal{D}_S| = 500$). Entity representations are learned via link prediction by optimizing the objective function in Eq. 9 using Algorithm 1, under two different settings: with and without alignment updates. Across ten runs, test f1-scores (mean and variance) are calculated and displayed in Figure 4, as a function of scale with increasing graph size.

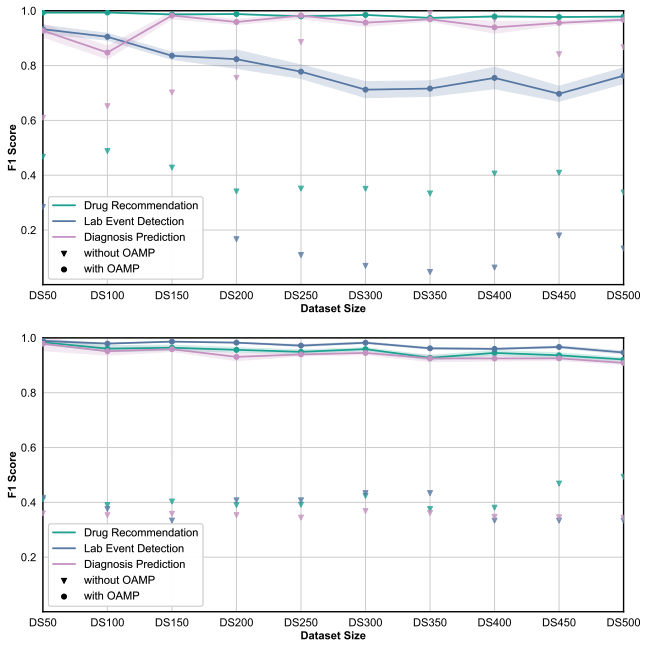


Fig. 4: Performance gaps with and without ontology-aligned message passing (OAMP), in terms of F1 Score for Link Prediction (Diagnosis/Lab/Medication) over MIMIC-III (top) and eICU (bottom) graphs with increasing size, from 50 visits ($|\mathcal{D}_S| = 50$) to 500 visits ($|\mathcal{D}_S| = 500$). Mean and variance values over 10 independent runs are displayed.

It can be observed that ontology-alignment has a crucial role in learning useful and task-specific embeddings by ensuring information exchange among nodes with disparate representations (i.e. fragmented codification). Over patient graphs from MIMIC III, without OAMP a single task dominates (i.e. diagnosis code prediction) while others continually deteriorate (i.e. lab observation and medication code prediction). This leads to a widening gap in concurrent performance due to fragmented and structure-less representations. Over patient graphs from eICU which is a multi-center dataset, without ontology alignment concurrent performances are severely impacted by the fragmentations in codification (f1-scores below 0.60). Remarkably, concurrent scores with OAMP uniformly converge around and above 0.80 points given the large quantity of alignments ($|\mathcal{G}_a| = 1,360$ and $|\mathcal{G}_a| = 13,999$). Concurrent prediction, as evident, requires capturing complex interactions among a diverse set of task-specific nodes (diagnosis, medication, and observation codes in this case) and without OAMP severe performance degradation is exhibited at scale.

V. Conclusions

In this work, we introduce a concurrent prediction framework that leverages ontology-aligned message passing to transfer both task-specific and structure-dependent information in a dual-space setting. Experiments on real-world datasets demonstrate that the model learns robust representations capable of supporting multiple downstream CVD prediction tasks, concurrently. The method consistently outperforms established baselines, achieving average precision, recall, and f1-scores of 0.98, 0.86, and 0.90. These results underscore the model's effectiveness in capturing heterogeneous clinical interactions and highlight the crucial role of implicit structural information in multi-task clinical prediction.

Declaration on Generative AI

Generative AI tools were used in a limited capacity. AI tools were used to assist with visual quality (color selection, font-size, etc in figures) and minor editing of the text (latex commands, etc), for better readability. The original implementation, experiment design, results, and any conclusions drawn are produced without employing Generative AI tools.

References

[1] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, and F. Yang, "Disease Prediction via Graph Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 818–826, Mar. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9122573

[2] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun, "GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 1126–1133, Jul. 2019, number: 01. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3905

[3] C. Mao, L. Yao, and Y. Luo, "Medgcn: Medication recommendation and lab test imputation via graph convolutional networks," *Journal of Biomedical Informatics*, vol. 127, p. 104000, 2022.

[4] G. P. Velo and P. Minuz, "Medication errors: prescribing faults and prescription errors," *British Journal of Clinical Pharmacology*, vol. 67, no. 6, pp. 624–628, Jun. 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2723200/

[5] L. Ferraz, P. G. Cotovio, and C. 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*, Jun. 2025. [Online]. Available: https://ceur-ws.org/Vol-4001/paper5.pdf

[6] J. Hao, C. Lei, V. Efthymiou, A. Quamar, F. Özcan, Y. Sun, and W. 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*, ser. KDD '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 2946–2954. [Online]. Available: https://doi.org/10.1145/3447548.3467138

[7] S. Mehryar, "Resolution-alignment-completion of tabular electronic health records via meta-path generative sampling," in *Proceedings of the 4th Table Representation Learning Workshop*. Vienna, Austria: Association for Computational Linguistics, Jul. 2025, pp. 200–207.

[8] X.-M. Zhang, L. Liang, L. Liu, and M.-J. Tang, "Graph Neural Networks and Their Current Applications in Bioinformatics," *Frontiers in Genetics*, vol. 12, Jul. 2021, publisher: Frontiers.

[9] X. Peng, G. Long, T. Shen, S. Wang, C. Zhang, A. Clarke, and C. Schlegel, "Mipo: mutual integration of patient journey and medical ontology for healthcare representation learning," in *2025 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2025, pp. 1–9.

[10] R. G. Iyer, Y. Bai, W. Wang, and Y. 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*, 2022, pp. 676–686.

[11] Y. Liu, J. Gu, C. Qiu, C. Yuan, and F. Gao, "Multi-hop knowledge graph embedding based on multi-task learning to assist personalized medication recommendation," in *Proceedings of the 2024 5th International Symposium on Artificial Intelligence for Medicine Science*, 2024, pp. 708–715.

[12] P. Jiang, C. Xiao, M. Jiang, P. Bhatia, T. Kass-Hout, J. Sun, and J. Han, "Reasoning-enhanced healthcare predictions with knowledge graph community retrieval," *arXiv preprint arXiv:2410.04585*, 2024.

[13] L. Ma, H. Han, J. Li, H. Shomer, H. Liu, X. Gao, and J. Tang, "Mixture of link predictors on graphs," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, ser. NIPS '24. Red Hook, NY, USA: Curran Associates Inc., 2025.

[14] S. Mehryar and R. Celebi, "Semantic annotation of tabular data for machine-to-machine interoperability via neuro-symbolic anchoring." in *SemTab@ ISWC*, 2023, pp. 61–71.

[15] T. H. Chan, G. Yin, K. Bae, and L. Yu, "Multi-task heterogeneous graph learning on electronic health records," *Neural Networks*, vol. 180, p. 106644, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608024005689

[16] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," *arXiv preprint arXiv:1412.6575*, 2014.

[17] R. Poulain and R. Beheshti, "Graph transformers on ehrs: Better representation improves downstream performance," in *The Twelfth International Conference on Learning Representations*, 2024.

[18] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, p. 160035, May 2016, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/sdata201635

[19] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eicu collaborative research database, a freely available multi-center database for critical care research," *Scientific data*, vol. 5, no. 1, pp. 1–13, 2018.

[20] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 135–144.

[21] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding," Mar. 2020, arXiv:2002.01680 [cs]. [Online]. Available: http://arxiv.org/abs/2002.01680

[22] Z. Liu, X. Li, H. Peng, L. He, and P. S. Yu, "Heterogeneous similarity graph neural network on electronic health records," in *2020 IEEE international conference on big data (big data)*. IEEE, 2020, pp. 1196–1205.

[23] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?" *Advances in neural information processing systems*, vol. 34, pp. 28 877–28 888, 2021.

[24] L. Rampášek, M. Galkin, V. P. Dwivedi, A. T. Luu, G. Wolf, and D. Beaini, "Recipe for a general, powerful, scalable graph transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 14 501–14 515, 2022.

[25] L. Bao, Y. Wang, X. Song, and T. Sun, "Hgcge: hyperbolic graph convolutional networks-based knowledge graph embedding for link prediction," *Knowledge and Information Systems*, vol. 67, no. 1, pp. 661–687, 2025.

[26] I. Chami, A. Wolf, D.-C. Juan, F. Sala, S. Ravi, and C. Ré, "Low-dimensional hyperbolic knowledge graph embeddings," *arXiv preprint arXiv:2005.00545*, 2020.

[27] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. 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*, 2017, pp. 787–795.

[28] Y. Guo, Y. Huang, Y. Wang, Y. Liu, S. Jing, T. Shan, Y. Miao, and B. Li, "Mdp-grl: Multi-disease prediction by graph-enabled representation learning," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2025.

[29] A. Roy and A. Saparov, "Transformers can learn connectivity in some graphs but not others," *arXiv preprint arXiv:2509.22343*, 2025.